

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



Peer-assessment within Dental Education

Tricio-Pesce, Jorge

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Peer-assessment within Dental Education

Jorge Alberto Tricio-Pesce

A thesis submitted for the degree of

Doctor of Philosophy

King's College London

2014

Intentional white page

Abstract

Background: Formative assessment with its feedback component has a powerful effect on students' learning.

Aim: This research aimed to appraise how teaching and assessment practices were organised in dental undergraduate teaching institutions to inform the development and piloting of a novel peer-assessment protocol for undergraduate dental students' pre-clinical competence and clinical performance. Subsequently, the protocol's utility as a framework for immediate dialogic peer-feedback to boost students' academic learning and achievement outcomes as well as their reflective skills, was evaluated.

Materials and methods: An initial review of the literature on peer-assessment together with a scrutiny of the King's College London Dental Institute undergraduate curriculum and assessment practices of 39 selected international dental teaching institutions was undertaken. This underpinned the development of a novel longitudinal, formative and structured peer-assessment protocol based on traditional Workplace-Based Assessment forms to be used as a framework for immediate peer-feedback and self-reflection. Subsequently, the protocol was piloted and later implemented in a larger trained sample to judge its utility towards fostering students' academic achievements and reflective skills. Thus, following a baseline quantitative reflection skills evaluation, volunteer students assessed their peers' pre-clinical competence (BDS year-2) and clinical performance (BDS year-5) across the whole academic year. Students' previous end-of-year examination and baseline reflection skills scores from the study and control groups (those who did and did not exercise the peer-assessment protocol, respectively) were compared to their current end-of-year examination marks and a second reflection skills evaluation score. Students' feedback narratives and their reasons to participate or not in the peer-assessment protocol, were also analysed.

Results and Discussion: Peer-assessment was only used by 19% of the surveyed dental schools. Both pre-clinical and clinical peer-assessment participating students demonstrated a reliable ability to identify those domains where they performed better as well as those which needed improvement. They also detected progress over time. Additionally, students' peer-assessment scores were positively correlated to their end-of-year examination. Inasmuch as students exercised ten or more peer-assessment encounters, they significantly increased their higher order thinking skills and final examination scores. Peer-feedback narratives from pre-clinical and clinical students differed in their content and sign, but corresponded in their specificity. Previous negative feedback experiences played a notable role in students deciding whether to participate or not.

Conclusions: Longitudinal (≥ 10 encounters), formative and structured peer-assessment and peer-feedback to encourage self-reflection of undergraduate dental students' pre-clinical and clinical skills, can reliably help them to improve their academic achievement and develop higher order thinking skills.

Table of Contents

Volume I

Abstract	3
Table of Contents	4
Table of Figures	8
Table of Tables.....	10
Index of Acronyms.....	16
Acknowledgements	18
General Introduction	20
Aims	22
Chapter 1	25
Literature Review	25
1.1 Introduction.....	25
1.2 Learning theories underpinning peer-assessment and peer-feedback.....	26
Behaviourism.....	27
Cognitivism	29
Constructivism.....	31
Experiential learning	33
Reflective practice learning	35
Andragogy	37
1.3 Learning Theories and the Curriculum.....	40
1.4 Assessment.....	42
1.4.1 Assessment purposes	43
1.4.2 Assessment principles	45
1.4.3 Is there a role for peers in assessment?	52
1.4.4 Reported benefits of peer-assessment	54
1.4.5 Psychometrics of peer-assessment	55
1.4.6 Students' perception of peer-assessment.....	58
1.4.7 Limitations of peer-assessment	61
1.4.8 Feedback.....	62
1.4.9 Feedback that works	70
1.4.10 Reflection and learning	77
1.4.11 Clinical workplace-based assessment and the role of peers.....	81
1.5 Literature review conclusions	90
Chapter 2	92

Teaching and Assessment Approaches in Dental Education	92
2.1 Introduction.....	92
2.2 Aim	95
2.3 Hypotheses.....	95
2.4 Materials and Methods.....	95
2.4.1 KCLDI teaching and assessment.....	96
2.4.2 Assessment practices and strategies online survey	96
2.5 Results.....	99
2.5.1 KCLDI teaching and assessment.....	99
2.5.2 Assessment practices and strategies online survey	103
2.6 Discussion.....	105
2.7 Conclusion	109
Chapter 3	110
Developing and Piloting a Structured Protocol for Dental Students’	
Prospective Peer-Assessment and Peer-Feedback	110
3.1 Introduction.....	110
3.2 Aim	111
3.3 Hypotheses:.....	112
3.4 Materials and Methods.....	112
3.4.1 Developing the instruments.....	112
3.4.2 Piloting and data collection	120
3.4.3 Data analysis.....	122
3.5 Results.....	123
3.6 Discussion.....	129
3.7 Conclusion	132
Chapter 4	134
Utility of a Method to Assess Dental Students’ Reflection Habits	134
4.1 Introduction.....	134
4.2 Aim	136
4.3 Hypotheses.....	136
4.4 Materials and Methods.....	136
4.4.1 The instrument.....	136
4.4.2 Participants and data collection	139
4.4.3 Data analysis.....	139
4.5 Results.....	140
4.5.1 Participants and data collection	140

4.5.2 Data analysis.....	141
4.6 Discussion.....	146
4.7 Conclusions.....	153
Chapter 5	154
Implementing a Structured Protocol for Dental Students' Prospective Peer-Assessment and Peer-Feedback.....	154
5.1 Introduction.....	154
5.2 Aim	156
5.3 Hypotheses.....	156
5.4 Materials and Methods.....	156
5.4.1 Instruments for peer-assessment and feedback	156
5.4.2 Instruments to assess academic achievement outcomes.....	157
5.4.3 Instrument for reflection skills assessment	157
5.4.4 Implementation and data collection	158
5.4.5 Students' stimulus	160
5.4.6 Data analysis.....	160
5.5 Results.....	164
5.5.1 Participants	164
5.5.2 Peer-assessment general data	164
5.5.3 Reflection Questionnaires general data.....	167
5.5.4 Reliability of peer-DOPS and peer-mCEX	167
5.5.5 Validity of peer-DOPS and peer-mCEX	168
5.5.6 Feasibility of peer-DOPS and peer-mCEX	173
5.5.7 Acceptability of peer-DOPS and peer-mCEX.....	174
5.5.8 Educational impact	175
5.5.9 Peer-assessment influence on students' reflective skills.....	182
5.6 Discussion.....	188
5.7 Conclusions.....	199
Chapter 6	201
Analysis of Dental Students' Peer-feedback Narratives from a Prospective Peer-Assessment Protocol	201
6.1 Introduction.....	201
6.2 Aim	202
6.3 Hypotheses.....	202
6.4 Materials and Methods.....	203
6.4.1 Quantitative coding of peer-feedback narratives.....	203

6.4.2 Qualitative thematic analysis of peer-feedback narratives	203
6.4.3 Focus groups on the reasons peer-assessment participating and non-participating students decided to take part or not in the protocol	203
6.5 Results.....	204
6.5.1 Quantitative coding of peer-feedback narratives.....	204
6.5.2 Qualitative thematic analysis of peer-feedback narratives.....	207
6.5.3 Focus groups on the reasons peer-assessment participating and non-participating students' decided to take part or not in the protocol	208
6.6 Discussion.....	210
6.7 Conclusion	217
General Discussion.....	219
Conclusion	224
Future Studies	225
References	226

Volume II

Appendix Book Contents	267
Appendix 1 List of Publications and Presentations	268
Appendix 2 Dental Students Peer-Assessment: a Prospective Pilot Study. European Journal of Dental Education (ID EJE-13-1242.R2 accepted)	271
Appendix 3 Dental Students' Reflective Habits: Is There a Relation with their Academic Achievements? European Journal of Dental Education (ID EJE-14-1382.R1 accepted)	295
Appendix 4 Ethical approval	318
Appendix 5 Mini-Clinical Evaluation Exercise (mini-CEX)	320
Appendix 6 Direct Observations of Procedural Skill (DOPS)	321
Appendix 7 Longitudinal Evaluation of Performance (LEP)	322
Appendix 8 Case-base Discussion (CbD).....	323
Appendix 9 Patient Assessment Questionnaire (PAQ).....	324
Appendix 10 Mini-Peer Assessment Tool (mini-PAT)	325
Appendix 11 Transcription of the KCLDI syllabus.....	326
Appendix 12 Pre-clinical peer-DOPS for BDS 2	495
Appendix 13 Clinical peer-DOPS for BDS 5	497
Appendix 14 Clinical peer-mCEX for BDS 5	499

Table of Figures

Figure A-1 Flowchart of the research chapters with their respective aims and study design.	24
Figure 1-1 Graphical representation of Vygotsky notion of the Zone of Actual Development (ZAD), where the learner can solve problems independently, and the Zone of Proximal Development (ZPD), where the learner can solve problems with support (Vygotsky, 1978).	32
Figure 1-2 Four stages of Kolb's experiential learning cycle (Kolb, 1976 p. 2) complemented with learning strategies as suggested by Knowles et al. (2012 p. 197).	33
Figure 1-3 Five stages of Schön iterative model of reflective practice following a given action [interpreted from Schön (1983)].	35
Figure 1-4 Graphical interpretation of learning theories related to peer-assessment through feedback.	40
Figure 1-5 Miller's pyramid for clinical assessment (Miller, 1990) and frequently used assessment methodologies (Kramer et al., 2009).	46
Figure 1-6 Graphical interpretation of the peer-feedback characteristics that could have a positive or negative influence. This could be used for peer-assessment training purposes.	76
Figure 2-1 Flowchart of the study's research chapters. The second current chapter is highlighted to help in providing an overall view.	92
Figure 2-2 "Preparing for Practice: Dental team learning outcomes for registration" four integrated and mutually supported domains (2012a p. 9).	93
Figure 2-3 Graphical representation of the percentage of teaching methodologies from all courses together.	99
Figure 2-4 Overall assessment framework within the four levels of Miller's Pyramid (Miller, 1990).	100
Figure 2-5 Summative, Formative and Hurdle assessment approaches mapped to the four levels of Miller's Pyramid (Miller, 1990).	101
Figure 2-6 Percentage of assessment sessions grouped according to Miller's pyramid of professional competence, per year.	102
Figure 3-1 Flowchart of the study's research chapters. The third chapter is highlighted to help in providing an overall view.	110
Figure 3-2 Initial header of the peer-assessment DOPS form which maintained the	

traditional norm-referenced assessment scale.	113
Figure 3-3 Header of the peer-assessment DOPS form containing now the criterion-referenced assessment scale.	114
Figure 3-4 Header of the first option of an educationally-referenced scale with written explanations of progress presented to students.	118
Figure 3-5 Header of the second option of an educationally-referenced scale with graphical explanations of progress presented to students.	119
Figure 3-6 Header of the six-point educationally referenced scale used in both pre- clinical and clinical peer-assessment instruments which asks the “observing” student to judge their peer’s ability over time.	120
Figure 3-7 Graphical representation of prospective peer-DOPS marks for every fortnightly assessment occasions (mean and standard deviation).	126
Figure 4-1 Flowchart of the study’s research chapters. The fourth chapter is highlighted to help in providing an overall view.	134
Figure 4-2 Mean score and \pm standard deviation for Habitual Action, Understanding, Reflection and Critical Reflection scales of the RQ, grouping all undergraduate students together (UG), clinical postgraduates (CPG) and PhD trainees.	143
Figure 5-1 Flowchart of the study’s research chapters. The fifth chapter is highlighted to help in providing an overall view.	154
Figure 5-2 Example of a BDS 5 student’s peer-feedback diary designed to help them recall their peer-feedback and agreed challenges to stimulate self- reflection (the name of the student has been hidden).	159
Figure 5-3 Overall mean peer-DOPS and peer-mCEX scores and standard deviations by session for BDS 2 and BDS 5 cohorts.	172
Figure 5-4 Graphical interpretation of Vygotsky ZAD (Zone of Actual Development) and ZPD (Zone of Proximal Development) (Vygotsky, 1978) according the results of the present chapter.	191
Figure 6-1 Flowchart of the study’s research chapters. The sixth chapter is highlighted to help in providing an overall view.	201

Table of Tables

Table 1-1 Pedagogical and Andragogical models comparison according to Knowles et al. (2012 pp. 298-299).	39
Table 1-2 Percentage of UK NSS higher education full time student's satisfaction ("mostly agree" and "definitely agree") to the three questions related to feedback (numbers 7, 8 and 9) at three different years (2007, 2010, 2013), for students from medicine and dentistry, education, as well as the overall score. Source: Higher Education Funding Council for England (2011 p.42) and Unistats (2014).	64
Table 2-1 Taxonomy used to approach the new 2012 GDC Learning Outcomes (General Dental Council, 2012a).....	94
Table 2-2 Template used for mapping the 2012 GDC learning outcomes with BDS teaching and assessing methodologies at KCLDI.	96
Table 2-3 Survey designed to assess how dental teaching institutions assessed each of the four GDC "Preparing for Practice" domains.	98
Table 2-4 KCLDI teaching methodologies by course year.....	99
Table 2-5 Learning outcomes range according to the keyword verbs from the proposed taxonomy for each of the four domains.	103
Table 2-6 Survey results showing the % of dental teaching institutions using the listed assessment methods for each of the four GDC "Preparing for Practice" domains.	104
Table 3-1 Pre-clinical peer-DOPS checking of the content validity. Each item to be assessed was present in one or more BDS 2 coursebooks learning outcomes. The GDC domains have been coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93).....	115
Table 3-2. Clinical peer-DOPS checking of the content validity. Each item to be peer-assessed was mapped to the BDS 5 coursebooks and the "Preparing for Practice" GDC learning outcomes (2012a). The GDC domains have been coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93).	116
Table 3-3. Peer-mCEX checking of the content validity. Each item to be peer-assessed was mapped to the BDS 5 coursebooks and the "Preparing for Practice" GDC learning outcomes (2012a). The GDC domains have been	

coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93).	117
Table 3-4 Anonymous questionnaire to study students' perceptions presented to both BDS 2 and BDS 5 groups at the end of the peer-assessment exercise.....	122
Table 3-5 Variance components estimates for each peer-assessment tools.....	124
Table 3-6 Mean and standard deviation (sd) of the peer-observation and feedback times (minutes), overall peer-assessment score (scale 1 to 6), and students' perception of the utility of giving and receiving feedback to improve future performance (scale 1 to 6), for each of the studied groups.	125
Table 3-7 Mean and standard deviation (sd) mark for each of the eleven pre-clinical peer-DOPS items showing the ability of pre-clinical BDS 2 participating students to identify differences ($F=3.94$, $p<0.0001$) in their peers' performance (57 completed forms).	125
Table 3-8 Mean and standard deviation (sd) mark for each of the thirteen clinical peer-DOPS items showing the ability of clinical BDS 5 participating students to identify differences ($F=6.55$, $p<0.0001$) in their peers' performance (104 completed forms).....	126
Table 3-9 Comparison of the BDS 2 and BDS 5 peer-assessment participating students high stakes end-of-year examination marks with those of their respective classmates who did not take part in the study.	127
Table 3-10 Selected BDS 2 and BDS 5 students' feedback comments and agreed challenges and actions, extracted from their peer-DOPS completed forms.....	128
Table 3-11 Anonymous BDS 2 ($n=10$) and BDS 5 ($n=16$) students' perceptions (%) of the peer-assessment protocol after using the peer-DOPS instruments for 6 and 7 occasion, respectively.	128
Table 4-1 Reflection Questionnaire from Kember et al. (2000) used to assess the reflective habits of students and trainees.....	138
Table 4-2 Number of participants by course of study, mean age and standard deviation (sd) and proportion of students with a previous university degree.	141
Table 4-3 Mean value for each of the four RQ scales of all subjects and ANOVA with post hoc Tukey's test results showing inter-scale mean difference and corresponding p-values.....	142
Table 4-4 Mean scores and standard deviations (sd) for each of the four scales of the RQ according to participants' academic course and study year.....	142
Table 4-5 Separate independent samples t-test results showing number, mean	

scores, standard deviations (sd) and p-values for each of the four RQ scales according to participants' gender, age group and whether they had a previous university degree.	144
Table 4-6 Mean value for each of the four RQ scales of all subjects and inter-scale Pearson correlations with the corresponding p-value.....	144
Table 4-7 Correlation and corresponding p-value between each RQ scale of BDS year 1-2-3-4-5 subjects and their academic performances in end-of-year examinations. For each examination, the cohort, number of subjects and the examination mean score of the sample as well as that from the whole class () are shown.....	145
Table 4-8 Selection of participants' comments, study cohort, age, previous degree and their scores in Habitual Action (HA), Understanding (U), Reflection (R), and Critical Reflection (CR).....	145
Table 5-1 Number of BDS 2 and BDS 5 students according to the number of peer-assessment encounters completed.	165
Table 5-2 Number of pre-clinical procedures BDS 2 students peer-assessed each other using peer-DOPS according to the perceived complexity assigned to each procedure.	165
Table 5-3 Number of clinical procedures BDS 5 students peer-assessed each other using peer-DOPS and peer-mCEX according to the perceived complexity assigned to each procedure.	166
Table 5-4 Summary statistics (mean (sd)) of various measures of BDS 2 and BDS 5 peer-assessment forms.....	166
Table 5-5 D study for different number of encounters for pre-clinical peer-DOPS, clinical peer-DOPS and peer-mCEX.....	167
Table 5-6 Variance components estimates for each peer-assessment form.	168
Table 5-7 The mean and standard deviation (sd) for the 11 peer-DOPS items for pre-clinical peer-assessment of BDS 2 students (N=516 observations).....	169
Table 5-8 The mean and standard deviation (sd) for the 13 peer-DOPS items for clinical peer-assessment of BDS 5 students (N=576 observations).....	169
Table 5-9 The mean and standard deviation (sd) for the 9 peer-mCEX items for clinical peer-assessment of BDS 5 students (N=77 observations).....	170
Table 5-10 Anonymous BDS 2 (N=26), BDS 5 (N=43) students' and Tutors' (N=9) perceptions (%) of the peer-assessment protocol, assessed at the end of the study.	170

Table 5-11 Correlation between peer-assessment scores and the mean of all end-of-the-year examinations for each of the three tested instruments.	171
Table 5-12 Selected anonymous students' and tutors' free text comments on the peer-assessment protocol.....	174
Table 5-13 BDS 2 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and effect size (d), for the students' average of all end-of-year examination marks as well as specifically for those examinations with the same format both at baseline and resultant observations, according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed. The latter examinations were written SAQs (short answer questions) and SNQs (short note questions), and online MCQs (multiple choice questions).	178
Table 5-14 BDS 5 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and effect size (d), for the students' average of all end-of-year examination marks as well as specifically for those examinations with the same format both at baseline and resultant observations, according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed. The latter examinations were essays on clinical scenarios, online SAQs (short answer questions), SNQs (short note questions) and online MCQs (multiple choice questions), and the OSCE (Objective Structured Clinical Examination).	179
Table 5-15 BDS 2 resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (independent-samples t-test), and effect size (d) when comparing all groups to the 114 non-participating students (No peer-assessment), for each of the students' end-of-year examination marks (SAQs (short answer questions), SNQs (short note questions), MCQ (multiple choice questions) according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.....	180
Table 5-16 BDS 5 resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (independent-samples t-test), and effect size (d) when comparing all groups to the 87 non-participating students (No peer-assessment), for each of the students' end-of-year examination marks (SAQs=short answer questions, SNQs=short note questions, MCQs=multiple	

choice questions, OSCE=Objective Structured Clinical Examination) according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.....	181
Table 5-17 BDS 2 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and the effect size (d) for each of the four scales of the RQ according to participants' academic year and whether they participated in the peer-assessment exercise.	184
Table 5-18 BDS 5 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and the effect size (d) for each of the four scales of the RQ according to participants' academic year and whether they participated in the peer-assessment exercise.	185
Table 5-19 BDS 2 resultant (2013) mean scores, standard deviations (sd), statistical significance (independent-samples t-test) of the difference (p), and the effect size (d), when compared to the 31 non-participating students (No peer-assessment), for each of the four scales of the RQ according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.	186
Table 5-20 BDS 5 resultant (2013) mean scores, standard deviations (sd), statistical significance (independent-samples t-test) of the difference (p), and the effect size (d), when compared to the 18 non-participating students (No peer-assessment), for each of the four scales of the RQ according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.	187
Table 6-1 Proportional quantitative coding of BDS 2 and BDS 5 peer-feedback narratives according to their content, sign and specificity.	205
Table 6-2 Themes identified in BDS 2 peer-feedback narratives presented in alphabetical order (not frequency) and their corresponding GDC domain (General Dental Council, 2012a).....	207
Table 6-3 Themes identified in BDS 5 peer-feedback narratives presented in alphabetical order (not frequency) and their corresponding GDC domain (General Dental Council, 2012a).....	208
Table 6-4 Themes identified during the focus groups as the reasons why students who followed the peer-assessment protocol decide to take part in it. Recurrent	

themes are shown in bold.	209
Table 6-5 Themes identified during the focus groups as the reasons why students decided not to take part in the peer-assessment study. Recurrent theme is shown in bold.....	209
Table GD-1 Aims and outcome summary for each research chapter of the thesis...	220

Index of Acronyms

ADEA	American Dental Education Association
ADEE	Association for Dental Education in Europe
ADS	Applied Dental Sciences
AMEE	Association for Medical Education in Europe
ANOVA	Analysis of Variance
ATSRP	Assessment of Transferable Skills in Removable Prosthodontics
BDS	Bachelor of Dental Surgery
BMD	Biomedical Sciences, Dentistry, Medicine and Natural & Mathematical Sciences Ethical Committee
CAL	Computer Assisted Learning
CbD	Case-base Discussion
CFD	Craniofacial Development
CNR	Case-Note Review
CODA	American Commission on Dental Accreditation
COPDEND	Committee of Postgraduate Dental Deans & Directors
CRE	Clinical Reasoning Examination
CSCD	Community Special Care Dentistry
CSR	Chart-Stimulated Recall
CTT	Classical Test Theory
DEP	Dental Evaluation of Performance
D-EP	Dental Evaluation of Performance
DNA	Deoxyribonucleic Acid
DOPS	Direct Observation of Procedural Skill
EMQ	Extended Matching Question
GABA	Gamma-Amino Butyric Acid
GDC	General Dental Council
GDP	General Dental Practice
GIC	Glass Ionomer Cement
GMC	General Medical Council
GRAS	Gröningen Reflection Ability Scale
GT	Generalizability Theory
HD	High Definition
IADR	International Association for Dental Research
ICA	In-Course Assessment
IOTN	Index of Treatment Needs
IRT	Item-Response Theory
ISCP	Intercollegiate Surgical Curriculum Programme
IV	Intra-Venous
KCL	King's College London
KCLDI	King's College London Dental Institute

LA	Local Anaesthesia
LEP	Longitudinal Evaluation of Performance
MCQ	Multiple-Choice Questions
MCT	Molecules Cells and Tissues
Mini-CEX	Mini-Clinical Evaluation Exercise
Mini-PAT	mini-Peer Assessment Tool
MOS	Minor Oral Surgery
MPS	Medical Protection Society
MSF	Multi-Source Feedback
NICE	National Institute for Health and Care Excellence
NSS	National Students Survey
OHP	Oral Health Promotion
OSCE	Objective Structured Clinical Examination
OSCT	Observed Structured Clinical Test
PAP	Peer-Assessment Protocol
PAQ	Patient Assessment Questionnaire
PBL	Problem-Based Learning
PDC	Primary Dental Care
Peer-DOPS	Peer-Direct Observation of Procedural Skill
Peer-mCEX	Peer-Mini-Clinical Evaluation Exercise
PPE	Personal Protective Equipment
RQ	Reflection Questionnaire
SAQ	Short Answer Question
SBA	Single Best Answer
SCOT	Structured Clinical Operative Test
SIG	Special Interest Group
SNQ	Short Note Questions
SOLO	Structure of the Observed Learning Outcome
SPRAT	Sheffield Peer Review Assessment Tool
SPSS	Statistical Package for the Social Sciences
TCA	Tricarboxylic Acid Cycle (Krebs cycle)
TEL	Technology Enhanced Learning
TMA	Tutor Mark Assessment
TMJ	Temporomandibular Joint
WPBA	Workplace-Based Assessment
ZAD	Zone of Actual Development
ZPD	Zone of Proximal Development

Acknowledgements

These three years of PhD work have been a wonderful and often overwhelming but positive learning experience. What is written in these pages is only part of it. Like in any other academic educational programme, it has involved learning and dealing with many other enriching individual and social learning experiences that have helped me in the endless duty of enhancing the talents and opportunities I have been blessed with. The most important learning experience has been to balance the professional responsibilities of following a challenging PhD with those memorable family life occasions that will never return. It is clear then now, that this has not been a solo race and therefore I am indebted to many.

Above all, I thank **God** for the above mentioned talents, His blessing, wisdom, favour, faithfulness, guidance and protection which otherwise would have made impossible for me to reach this endpoint.

My deepest thanks to:

Professor Mark Woolford, my first supervisor, who from the very first day I visited the Dental Institute, even before starting this academic challenge, trusted me. His initial inspiration became later a permanent and paramount encouragement and support not only for the successful conclusion of the PhD, but also to complement my programme with the best educational training opportunities available in the UK and overseas. I am grateful to him for being *always* available to listen to me in those early morning talks and especially to help me in professional as well as personal matters. His wise advices to turn towards my family during those difficult times of the PhD are much appreciated.

Dr Michael Escudier, my second supervisor, for his trust and knowledgeable initial advices on starting to write up the thesis from the very beginning which proved correct. His throughout help to critically understand and interpret the sometimes conflicting and entangled results of the current clinical studies, were crucial. Furthermore, his impressive analytical and astute experience on writing scientific papers and on expressing the intended ideas correctly were essential to get the current chapter's studies speedily accepted for publication.

Professor Agamemnon Grigoriadis, my postgraduate coordinator, for his advices, prompt availability for a relax chat, and his encouraging formative evaluations of progress during the PhD.

Professor Stephen Dunne, Professor Avijit Banerjee and Dr Lyndon Cabot, for trusting my research proposal and allowing and facilitating me to approach years 2 to 5 BDS students and invite them to participate in the current study protocol. Similarly, I am in debt to all

these student's tutors for relying on me and permitting my presence during every single pre-clinical and clinical session whilst the study was implemented. They are many, but I cannot fail to mention **Dr Helena Lewis-Greene, Dr Sofia Zacharaki, Dr Michael Thomas, Dr Lakis Georghiou, Dr George Aristidou, Dr Dipesh Patel, Dr Charles Szentesi, and Dr Paul Duvoisin.**

Professor Tim Newton for his help and support on the reflection studies of this thesis.

All **King's dental students** who voluntarily decided to take part in this study and enthusiastically followed the research protocol during the whole academic year. There is no way this study would have been possible without their invaluable contribution and appreciated comments and suggestions.

Mr Manoharan Andiappan for his teachings and patience in the sometimes difficult analysis of the huge amount of data.

The Dental Institute Central Office staff members for their always friendly smile and help in all those big and small details that made my PhD experience easier and enjoyable. **Mrs Diane Brook** and **Mrs Margaret Whatley** deserve a special line in these acknowledgements.

My friends at my Homeland University of los Andes in Santiago, Chile. They gave me their unconditional support from the very beginning planning of this UK adventure. My deep appreciation and debt of gratitude goes to all the UAndes family and especially to the former Principal **Professor Orlando Poblete**, former vice-Principal **Professor Jaime Arancibia**, and to the current Principal **Professor José Antonio Guzmán**. I am also in debt to my friends and colleagues at the Dental School Council for endorsing this PhD challenge as if it was their own. **José Antonio, Francisco** and **Cristian**, I shall never forget your support.

My practice associate, colleague, and life-long friend **David Rosenberg** without whose support and encouragement, throughout the process, this PhD would not have been possible: I know the practice and patients have been in "good hands".

Finally, the whole dream of this PhD would have not been possible without the courage of my wife who never doubted to embark the whole family in this venture. Her faith and wise vision of a better future life together guided us all during the pleasant and difficult moments we have spent in the UK. Her unconditional support and sublime care of our boys during the long working months gave me the peace of mind that allowed my dedication and focus on the PhD. To my children, thank you for taking care of your daddy, for helping me to keep good track of the word counts and celebrating the finished chapters, and especially for keeping the house in silence while I was working.

My beloved family, **Carolina, Ignacio** and **Agustín** this PhD thesis is dedicated to you...

General Introduction

Traditionally, the model of dental education has been based on the early teaching of the biologic and clinical foundations which underpin all areas of performance, followed by the increasing practice of technical competencies in a laboratory environment and pre-clinical simulation, organised in different parts of the curriculum. Subsequently, once students are shown to be safe, they move from the laboratory into the work environment. That is the clinical environment with real patients, in a model that integrates instructive teaching with real clinical practice. Thus, a varying number of novice learners are guided by an “expert clinical practitioner”, the instructor, in the tradition of the artisan-apprentice training model (Hendricson and Cohen, 1999).

However, this embodies an ever present challenge for clinical instructors as they are required to simultaneously engage in high-quality patient care by ensuring a right diagnosis and treatment, and assess the clinical skills and reasoning of learners (Bowen, 2006). The former depends on factors that most of the time are outside the instructors’ control such as the patient type, multiple clinical situations and time sensitivity (Hoffman and Donaldson, 2004). This setting does not help in identifying students’ specific difficulties (Audétat *et al.*, 2013) and explains why they are seldom observed and provided with feedback during workplace performance (Norcini and Burch, 2007). Furthermore, the imparted knowledge takes supremacy over the way this knowledge is given (Clarke, 2011), and information passes from the instructor to the novice learner (Harden *et al.*, 1984). In this model, how can we ensure that future professionals, as a competent workforce, keep on learning? (Southgate and van der Vleuten, 2014). If students get dependant on their instructors’ feedback they will hardly develop the needed capacity to assess their own work, and that of others. Consequently, a more student-centred model has been promoted which focuses on the active process of learning (Nicol and Macfarlane-Dick, 2006) and on what students do to achieve their knowledge (Harden and Crosby, 2000), where collaborative and social peer-assessment are encouraged (Chickering and Gamson, 1999) in order to increase the opportunities and frequency of students’ feedback to develop their understanding, learning and reflection (Dochy *et al.*, 1999; Boud and Falchikov, 2007; Biggs and Tang, 2011).

However, as will be discussed in the subsequent chapters of this thesis, both self- and peer-assessment studies in dental education are scarce and those few available are primarily focused on the marking process and not on the learning effects of the

experience. Chapter 1 explores the literature on the learning theories underpinning peer-assessment and peer-feedback. The review goes into greater depth on assessment purposes and principles in addition to the role of peers in it, in order to clarify the potential educational benefits as well as the limitations of peer-assessment reported in different areas of education. Peer-feedback and self-reflection, as the educational components of peer-assessment, are also analysed in the search for an underpinning model that will allow us to develop a peer workplace-based assessment protocol to be implemented in undergraduate programmes.

As a result of the limited evidence of peer-assessment in dental education, I examine the teaching and assessment practices at King's College London Dental Institute (KCLDI) as a foundation knowledge of teaching and assessment approaches. Subsequently in Chapter 2, I report an international survey of dental teaching institutions' assessment practices, especially probing the extent of peer-assessment.

Chapter 3 amalgamates both previous chapters' results to develop and pilot a longitudinal peers' protocol of observation, scaffolding, assessment and feedback. The consequent challenge, described in Chapter 4, was to probe a valid and reliable method to assess dental students' reflection. Chapter 5 depicts the pivotal study of implementing the developed and piloted peer-assessment protocol in combination with the reflection evaluation method, in a larger sample throughout the whole academic year to trial its educational impact. Finally, in order to better understand how the protocol could have influenced students' skills, Chapter 6 analyses students' written peer-feedback narratives. In the hope of further developing the protocol, this Chapter also describes the motives that drove students to take part or not in the volunteer peer-assessment exercise.

Aims

Adopting aspects of Social Constructivist theory and Reflective Practice Learning as a theoretical framework (Cook *et al.*, 2007), this PhD research aims to appraise how teaching and especially assessment practices are organised in dental undergraduate teaching institutions to inform the development, piloting and utility evaluation of a new peer-assessment protocol of undergraduate dental students' preclinical-competence and clinical-performance. This framework was used for immediate dialogic peer-feedback in order to attempt to boost students' academic learning and achievement outcomes as well as their reflective skills.

In order to meet the above primary aim, each chapter of the project contributed with secondary aims (Figure A-1). All these received full ethical approval from King's College London Biomedical Sciences, Dentistry, Medicine and Natural & Mathematical Sciences Ethical Committee (reference number BDM/11/12-21) (Appendix 4).

Grouped by chapter, secondary aims are:

Chapter 2

- To investigate the different teaching and assessment methodologies used in dental education using King's College London Dental Institute (KCLDI) as true model.
- To investigate assessment practices and strategies used by selected dental teaching institutions around the world to measure students' progress particularly in relation to the use of peer-assessment.

Chapter 3

- To develop and pilot a structured protocol of formative, prospective peer-assessment of undergraduate pre-clinical and clinical dental students' skills in all four clinical, communication, professionalism, and management and leadership General Dental Council (GDC) domains, which would be used as an informed framework for the provision of immediate peer-feedback.

Chapter 4

- To assess the usefulness, in terms of reliability and construct validity of a self-reported Reflection Questionnaire (RQ) as a method of assessing dental students' and postgraduate trainees' reflection habits.

Chapter 5

- To evaluate the reliability, validity, feasibility, acceptability and educational impact of the same structured protocol (Chapter 3) of formative, prospective peer-assessment of undergraduate pre-clinical and clinical dental students' skills as a framework for the provision of immediate peer-feedback. Further, the influence of the peer-assessment protocol on the students' reflective skills was also examined.

Chapter 6

- To quantitatively and qualitatively analyse students' written peer-feedback narratives from the peer-assessment exercise, as well the reasons students decided to take part in, or not take part in the peer-assessment protocol.

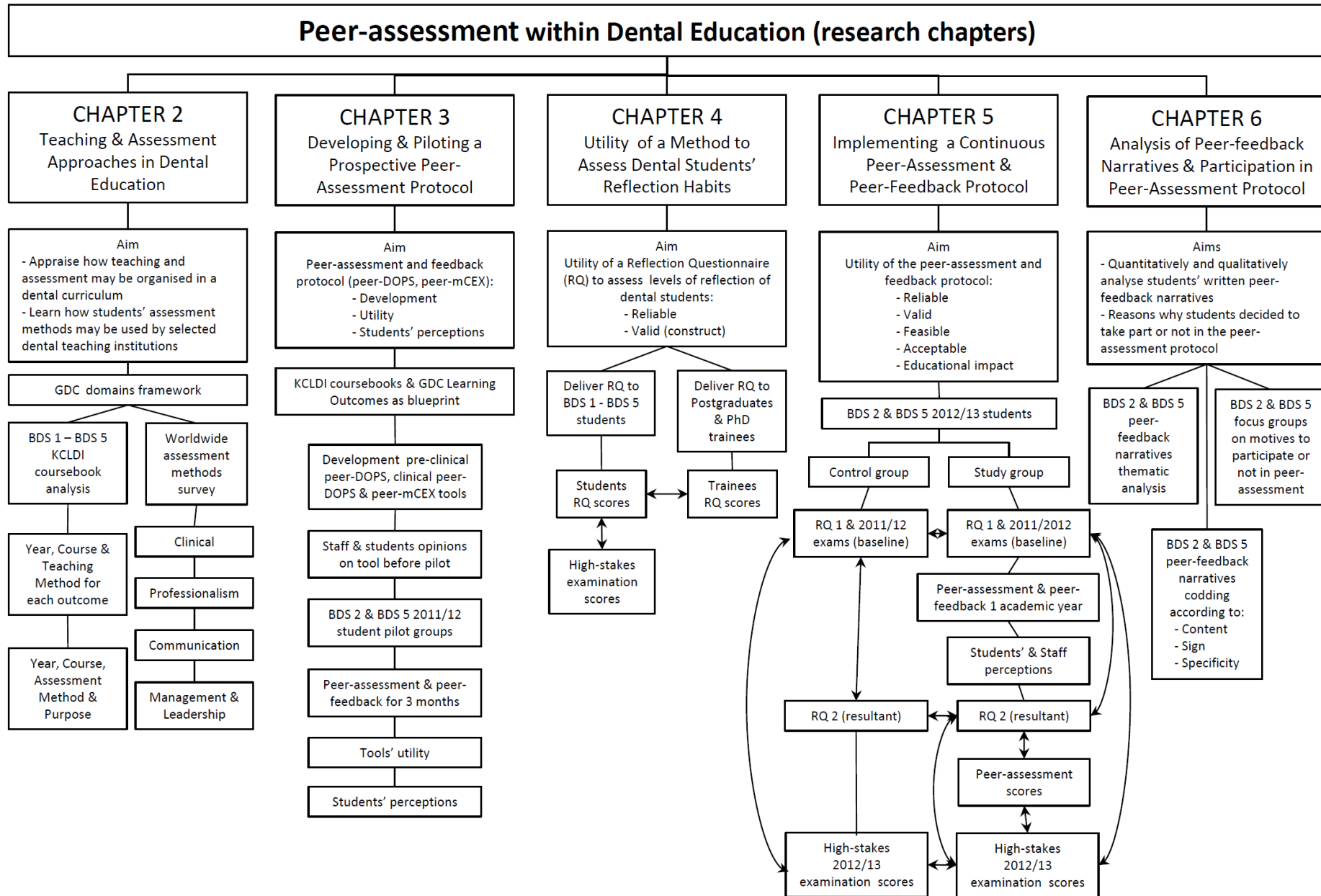


Figure A-1 Flowchart of the research chapters with their respective aims and study design.

Chapter 1

Literature Review

1.1 Introduction

Learning is the process of bringing together not only knowledge, but emotional and environmental influences and experiences for acquiring, enhancing, or making changes in one's knowledge, skills, values, and views (Ormrod, 2011). It is also about how we perceive and understand the world, about making meaning (Marton and Booth, 1997). Further, the process of learning has been described as being at the heart of life itself (Jarvis, 1987), and one of the most important activities in which humans engage. It is the core of every educational process (Pritchard, 2013), and we as teachers must understand and be aware of this big responsibility.

Consequently, our role as educators must be assumed as a comprehensive, intellectually challenging and facilitated onward process (Oliver *et al.*, 2008; Jahangiri and Mucciolo, 2011) that enables students to learn how to acquire knowledge, how to make the right choices, how to integrate new developments into existing knowledge and how to apply this new knowledge (Kersten, 1997). According to Harden and Laidlaw (2013), most learners in the healthcare professions are very capable and should have little difficulty in achieving the required learning outcomes. However, they will need help, and professional teachers cannot operate using a cookbook approach blindly following guidelines (Harden *et al.*, 1999). Mounting criticism of the quality and efficiency of teaching in higher education (Light *et al.*, 2009), creates the need for tutors to have knowledge about the underlying learning principals as well as the best educational evidence to inform their decision process to facilitate students learning. After all, the aim of teaching is to make student learning possible (Ramsden, 2003), and unfortunately some academics teach students without having much formal knowledge of how students learn (Fry *et al.*, 2009). However, it is understandable, especially for newcomers to the field of education, that different and new pedagogical practices, educational philosophies and conceptual frameworks, can create some initial confusion (Swanwick, 2011).

Accordingly, teaching skills have become increasingly important in view of the new way of creating and using knowledge which focuses on solving problems (Ramsden, 2003), that is, teachers need to help students to develop the ability to employ knowledge

during their practice in such a manner as to devise, choose and apply appropriate responses to unexpected and sometimes complex problems of the present and the future (Light *et al.*, 2009; Barber *et al.*, 2013). This is not an easy task especially for busy clinical tutors. However, as discussed below, encouraging students' reflection might facilitate a problem solving approach to practice.

For this reason it was considered important to start this review of the literature approaching the learning theories which underpin adult learning and especially peer-assessment and peer-feedback.

1.2 Learning theories underpinning peer-assessment and peer-feedback

The nature of learning, what happens when we learn, how it happens and how to influence the process, has been the matter of study for centuries (Biggs and Tang, 2011). Explanations of the “what” and “how” when we learn, constitute learning or educational theories. These are based on evidence or long-term successful experience and work as useful guiding principles to inform practice (Kaufman, 2003), but not as a mathematical formula (Dennick, 2012). Until recently, psychologists were concerned with developing the “One Grand Theory of Learning” that covered all learning. This notion is now dead and there are several learning theories that explain how humans learn (Biggs and Tang, 2011).

It was deemed relevant to this research project to analyse the main learning theories, as they will be used as a framework to explain the results of the learning interventions described in the successive chapters.

There are too many specific theories of learning for them all to be reported here. However, most of the suggested theories in which peer-assessment and feedback as a learning activity is grounded, fall into three philosophical frameworks (Falchikov and Goldfinch, 2000). These are cognitivism (information processing, including insight, memory, perception, and meta-cognition), constructivism (construct meaning from experience), and andragogy (self-actualisation and continuing personal development) (Merriam *et al.*, 2007).

Although ‘behaviourism’ (shaping behaviour through rewards) has been strongly criticised (Chomsky, 1975; Jarvis, 1987), it will be the first theory to be described here,

as feedback, the learning component of peer-assessment, originates in behaviourism (Thorndike, 1911). Further, the intended reflection stimulus of the peer-assessment protocol implemented in this study and subsequently discussed, also suggests the inclusion and discussion of reflective theories of learning (reflection leads to action and then change).

Behaviourism

Behaviourism as a theory began with the work of Thorndike on animal intelligence (1911) and those of Watson (1913) in animal behaviour and child education. The theory centres on learning as the acquisition of new behaviour through ‘conditioning’. It disregards how individuals learn, or the cognitive, or mental, activity involved while this is taking place (Pritchard, 2013). Thus, the attention is focused on the rewarding (reinforcement), of a desired behaviour that occurs as a response of a particular stimulus from the surrounding environment (Knowles *et al.*, 2012). There are two types of conditioning that are described as viable explanations of the way animals and humans can be taught to do something (Pritchard, 2013).

First, the “classical conditioning”, where the desired behaviour to be learnt becomes a ‘reflex’ response to a specific stimulus. A well-known example of this is the work of Ivan Pavlov (1927) who conditioned dogs to salivate (reflex) at the sound of a bell (stimulus), as this was an indication that food was imminent (conditioned stimulus) (Merriam *et al.*, 2007; Pritchard, 2013).

Second, the “operant conditioning”, where the desired behaviour to be learnt is reinforced by a reward or a punishment. Again, known examples of this are the studies of probably the most famous behaviourist, the American psychologist Burrhus Frederic Skinner. He studied the response of rats and pigeons to earn a pellet of food (reward) when they pressed a lever. As this reward continued for the repetition of the action, the animal ‘learnt’ that to be fed it must press the lever (Skinner, 1958). Upon the basis of these and other experiments, educationalists have tried to make comparisons between animal and human learning (Wang, 2012).

Therefore, in behaviourism the environment is what shapes students’ behaviours. Consequently, it is the role of the teacher to provide, on more than one occasion, the required environments that prompt the desired behaviours to be learnt (Knowles *et al.*, 2012). Subsequently, the learning programme starts from the student’s initial

knowledge, moving forward in small broken steps which are easy to master (Cross, 1986 p. 232), and, as Skinner emphasised, continually reinforcing success rather than on punishing failure (to keep motivation high), through the provision of immediate feedback (Thorndike, 1911; Skinner, 1954). As a result of the rewards, those students who benefit the most are the less well motivated, the anxious, uncertain and the failing ones. On the other side, bright students can find this programmes unsatisfying and even boring (Pritchard, 2013). While behaviourist positions of learning are sometimes at odds with the current student-centred focus of adult education (Knowles *et al.*, 2012), they are frequently the foundation for one of the largest segments of adult education, such as job and skills training, and self-instructional packages (Cross, 1986; Knowles *et al.*, 2012 p. 125).

In health education, behaviourism is the basis of many objectives and competency-based curricula, and skills development programmes (e.g. tasks under time conditions) (Merriam, 2001; Taylor and Hamdy, 2013). Modelling behaviour through rewards and the influential factor of feedback reinforcement to learn, is common practice and central in health education (Mann, 2011). For example, a behaviourist learning approach would be the “skills and drills” of resuscitation training, which looks at making these ‘behaviours’ automatic (Ker and Bradley, 2011 p. 169).

However, teaching based on behaviourism is usually criticised as it results in learning that promotes standardisation of the outcomes (Taylor and Hamdy, 2013) and rote learning without understanding (Pritchard, 2013). Further, it is posited that animals learn via reflexes and behaviour modification, whilst humans learn through reflection (Wang and King, 2007). The acquisition of these higher order skills such as learning of language, argued Chomsky (1975), are not explained by behaviourist experiments.

Challenging Skinner’s (1954) behaviourist view, Jackson (2009) labels it as a simplistic approach to learning that treats individuals as subjects who are completely dependent on the environment that surrounds them. Further, the behavioural definition of learning, as the product (behaviour modification) of a particular process, is contended by Jarvis (1987), by stating that learning is both a cognitive process and a product. The present thesis will study both the process and product of the peer-assessment protocol, and consequently, the next learning theories will concentrate on the processes that occur in the person who is learning.

Cognitivism

Cognitive learning theories study human thinking to promote learning and started to supersede behavioural psychology in the late 1950s and 1960s, when psychologists and educators began to stress more complex cognitive processes such as thinking, problem solving, language, concept formation and information processing (Ertmer and Newby, 1993). Thus, cognitivism focuses learning not in behaviour but in the mental and psychological processes of the mind, such as perception, insight, information processing (Taylor and Hamdy, 2013), and memory (Regehr and Norman, 1996; Roediger and Karpicke, 2006; Brown *et al.*, 2014).

Cognitivists, such as Piaget (1952) and Bruner (1966) have had great impact on thinking about learning (Knowles *et al.*, 2012). They posit that learners develop new ideas, constructs, hypotheses and decisions based on their interaction with the world and their prior knowledge that exists in their minds, as an internal purposive mental process (Merriam *et al.*, 2007). Thus, the focus is on how information is received, stored and retrieved by the mind (Ertmer and Newby, 1993). Learning is assimilated when the new experience fits into existing structures and adds to the body of examples. In case the experience does not fit the existing structures, learning is accommodated and the structure must be changed in order to incorporate the new knowledge into a cognitive structure that gives meaning and organisation to the knowledge (Ker and Bradley, 2011).

This newly learnt knowledge is then stored in the long-term memory (Kirschner *et al.*, 2006) in the form of schemas which categorise elements of information according to the manner in which they will be used (Sweller *et al.*, 1998). The development of networks between schemas and the use of keywords allow the learners to link previously related experiences to the new one. Therefore, there is an imperative need for the learner to be presented with meaningful teaching experiences that at the same time are linked to pre-existing ones (previous knowledge) (Merriam *et al.*, 2007) as only if the novel information is useful will it trigger particular responses to learn (Regehr and Norman, 1996). This is usually performed using analogies and metaphors to help the learner conceptualise, organise and retain the new information (Ertmer and Newby, 1993).

Thus, cognitive psychology has helped in the understanding of how knowledge is organised and stored in the brain, how memory works, and how learners can be helped to make meaning of their experiences (Merriam *et al.*, 2007). As behaviourism,

cognitivism emphasises the role of the environment in facilitating learning. Likewise, the role of practice with feedback is also stressed in both approaches: behaviourism uses it for reinforcement; cognitivism for guiding and supporting accurate mental connections (Kirschner *et al.*, 2006). However, in cognitivism the main focus is on the mental activities of the learner, and as such is more appropriate for explaining complex forms of learning (reasoning, problem solving, information processing) than are those of behaviourism (Ertmer and Newby, 1993).

In health education, the concepts of cognitivism have clarified the important processes of clinical reasoning, decision making and problem solving (Kirschner *et al.*, 2006; Merriam *et al.*, 2007; Mann, 2011; Brown *et al.*, 2014). Furthermore, the instruction method of Problem-Based Learning (PBL) is strongly influenced by cognitive psychology (Norman and Schmidt, 1992). Implementing cognitivism, requires the tutor to establish students' preconceptions in order to present them with the new knowledge in context and at a suitable level of complexity (Knowles *et al.*, 2012). It is also common practice in a cognitivist approach to teaching to create and facilitate access to experiences in order to facilitate learning (Taylor and Hamdy, 2013). Consequently, students are presented with a cognitive conflict in a given experience, drawing attention to the discrepancy between what they expect according to their current knowledge, and the real experience of the event (new knowledge). This is achieved by asking them questions and establishing a dialogic feedback, in which they get to know that the new 'ideas' are better than their previous knowledge (Ker and Bradley, 2011). Feedback is then fundamental and acts as a learning catalyst highlighting the gap between students' actual knowledge and the level they need (Kluger and DeNisi, 1996; Archer, 2010).

Cognitivism as that from Piaget (1952) and Bruner (1966), has been criticised by the overemphasis on cognitive skills at the expense of emotional development; on knowledge attainment ignoring concept formation or invention; and on individual over common motives (Knowles *et al.*, 2012). This very last issue has been also a matter of discussion in the medical education literature. Swanwick (2005) has argued that a cognitivist approach in which the mind is treated as functioning independently of the learners' social context, is inadequate to explain the full process of clinical learning. This view has been increasingly supported by other healthcare researchers (van der Zwet *et al.*, 2011; Watling *et al.*, 2013) and the present study investigator, especially in the context of clinical workplace learning.

Constructivism

The primary idea of constructivism is that learners actively construct their own knowledge from a concrete experience (Dennick, 2012), in a cumulative nature- and context-bound, that is, based on the activation of what they already know, have experienced and feel, without separating it from the context in which it is used (Kaufman, 2010; Knowles *et al.*, 2012). Further, learners make personal judgments of the knowledge meaning (Merriam *et al.*, 2007) and so about when and how to modify it (Kaufman, 2010). Thus, constructivists believe that we learn best when we actively construct our own understanding from experiences (Pritchard, 2013).

In contrast to cognitivism, the constructivist perspective does not see the teacher as the transmitter of what needs to be acquired, but as a guide who facilitates the construction of the learner's own new knowledge, based upon experiences and cultural factors (Kaufman, 2010).

Social constructivism (Jarvis, 1987), based on the work of Lev Semyonovich Vygotsky (1978), goes one step further, stressing the importance of social interaction in the learning process (Morris and Blaney, 2011). Thus, understanding and learning from experiences is more successful when the learner engages in dialogue with a more knowledgeable 'other' - a teacher, a peer or others in their environment - than when doing it alone (Vygotsky, 1978). The focus is then on the way the student's community supports learning (Taylor and Hamdy, 2013). It is important to note, however, that a more knowledgeable 'other' does not imply someone older (e.g. parent) or in a position of responsibility for learning (e.g. tutor). It is perfectly possible for a peer, friend or even a younger sibling to take this role, especially in casual and informal situations (Pritchard, 2013).

According to Vygotsky (1978), learning would occur by awakening developmental processes that only operate when the learner is interacting with others, stressing the importance of language or 'shared talk' (feedback). Thus, through dialogue ideas are considered, shared and developed (Pritchard, 2013).

Further, Vygotsky introduced a fundamental new approach that has impacted on practice over the last 20 years or so, as the novel learning needs to be matched to the learner's developmental stage through the *Zone of Proximal Development* (ZPD) (Morris and Blaney, 2011). The ZPD is a theoretical space into which the learner can

move with the support of a more knowledgeable other (Vygotsky, 1978 p. 86; Pritchard, 2013), in contrast to the *Zone of Actual Development* (ZAD), where they can work unassisted (Figure 1-1). Thus, learning is what takes place in the ZPD, which is just above the understanding level of the learner (the ZAD), where she or he can work but only with guidance, assistance, support and coaching (feedback) (Morris and Blaney, 2011). The term ‘scaffolding’ was later introduced by Wood *et al.* (1976) to explain this process of help. Subsequently, at the appropriate time, the scaffold is gradually removed to encourage independence (Kaufman, 2010; Ker and Bradley, 2011). A frequently overlooked observation in Vygotsky’s work is his clear distinction between ‘learning’ and ‘development’. Whilst the latter, as the ZPD, is the *maturing* of the learner’s psychological functions, and not only what students can do with support. Thus, ‘learning’ involves the acquisition of new skills, without changes in the available psychological functions (Vygotsky, 1978 pp. 79-91; Black and Wiliam, 2009).

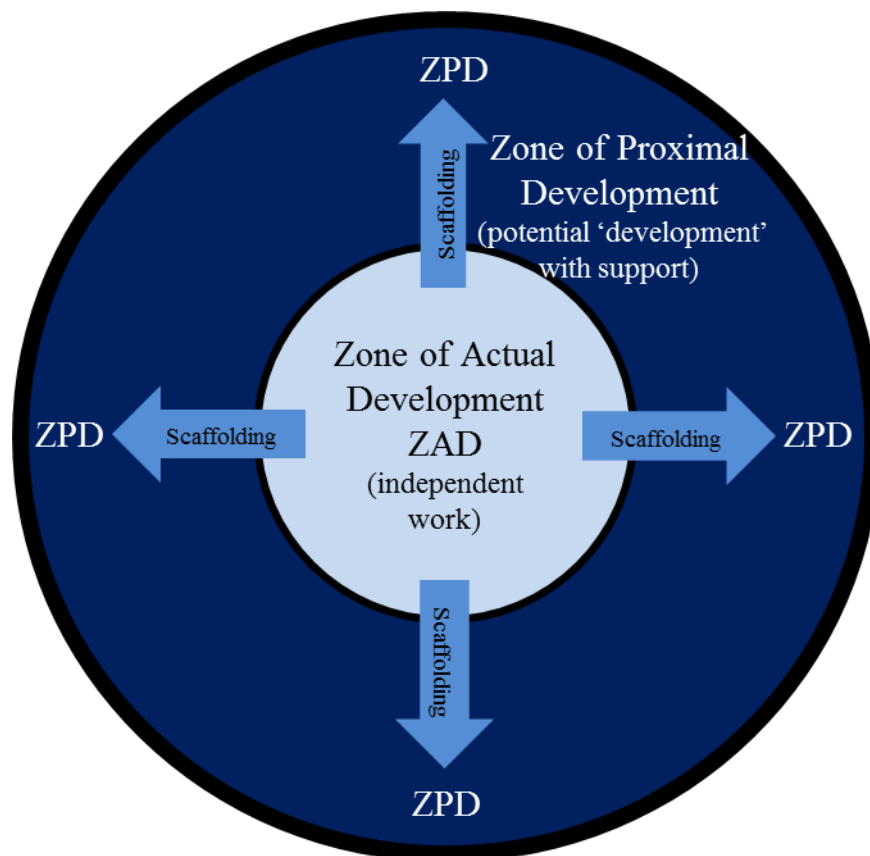


Figure 1-1 Graphical representation of Vygotsky notion of the Zone of Actual Development (ZAD), where the learner can solve problems independently, and the Zone of Proximal Development (ZPD), where the learner can solve problems with support (Vygotsky, 1978).

In a constructivist approach, effective teaching needs to identify the students' current competence level (the ZAD) and then offer help and set challenges that are somewhere ahead (the ZPD) of her or his competence (Kaufman, 2010). Despite what social constructivism having helped in providing models to support students' learning in dental education (Chadwick *et al.*, 2002; Moore and Kain, 2011; Bridges *et al.*, 2014; Postma and White, 2014), it is limited in explaining the complexities of workbased learning (Lave and Wenger, 1991; Wenger, 1998; Morris and Blaney, 2011).

Further, constructivism in general does not provide the mechanism for how learning takes place. Consequently, experiential learning, as developed by David Kolb (1984) will be explored next to explain how experience can be transformed into knowledge, skills and attitudes.

Experiential learning

Experiential learning theory is a holistic model of learning that postulates that “knowledge results from the combination of grasping experience and transforming it” (Kolb, 1984 p. 41). Experiential learning is therefore the ability to learn by adapting ideas as a result of experiences (Barley, 2012). For David Kolb, “learning is not so much the acquisition or transmission of content as the interaction between content and experience, whereby each transforms the other” (Knowles *et al.*, 2012 p.195).

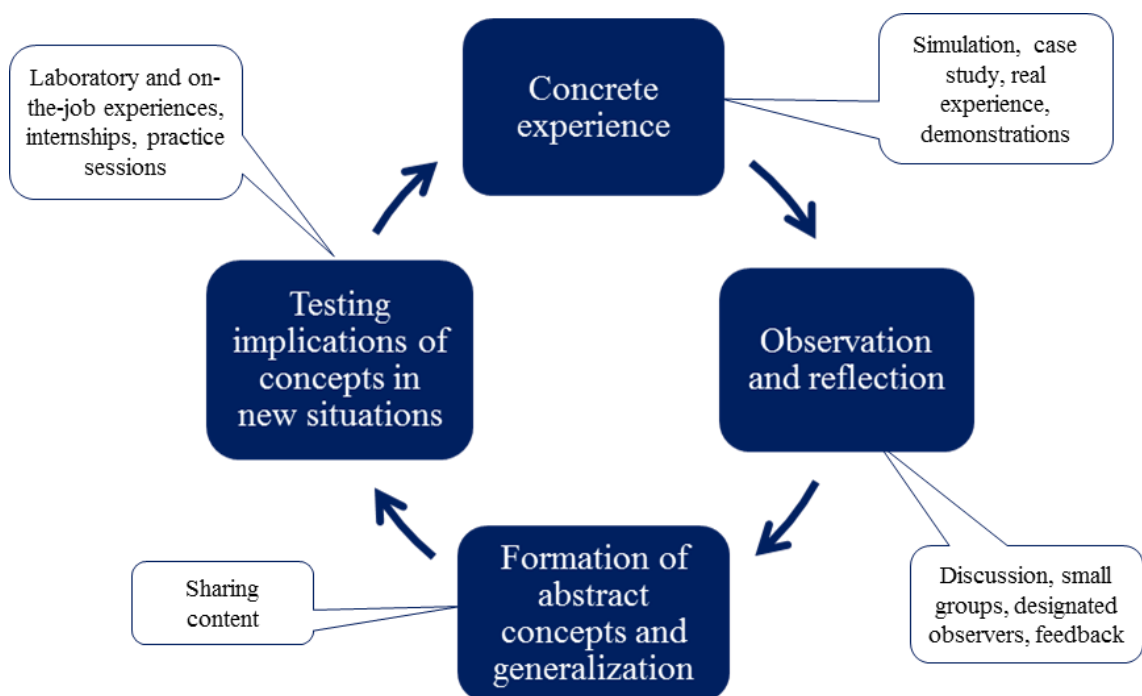


Figure 1-2 Four stages of Kolb's experiential learning cycle (Kolb, 1976 p. 2) complemented with learning strategies as suggested by Knowles *et al.* (2012 p. 197).

Kolb's learning cycle (Figure 1-2), stressing the critical importance of experience to learn, helps to explain the gap between students' understanding something or having just the knowledge, and having the competence to practice it. This might rationalise the strong component of students' practice at skills laboratories frequently found in dental curriculums. Experiential learning is therefore 'active' by definition (Light *et al.*, 2009 p. 55), and probably due to this is the most widespread theory of learning from experience (Fry *et al.*, 2009 p. 15).

The core of the theory conceives learning as a four stage continuous cycle. Immediate *concrete experience* (i) is the basis for *observation and reflection* (ii). These observations are assimilated into theory, *formation of abstract concepts and generalizations* (iii), from which new implications for action can be deduced, *testing implications of concepts in new situations* (iv), and these implications or hypotheses then serve as guides in acting to create new experiences (Kolb, 1976 p. 2). Despite in practice the learning process starting with an experience, the experiential learning can begin at any of the four stages (Kaufman, 2010). Kolb's cycle is a practical model for experiential learning practice (Knowles *et al.*, 2012), and though it emphasises individual learning, has been widely accepted in healthcare education (Mann, 2011). Probably the best accepted aspects of Kolb's theory in medical education are the use of experiences to test new knowledge, the opportunity to provide feedback to change students practices (Kaufman and Mann, 2010), and the increase motivation for learning and reflection that experiential learning produces in students (Ker and Bradley, 2011). It is important to be aware that in practice, students may get stuck somewhere in the cycle, fail to progress or jump about stages (Fry *et al.*, 2009).

Kolb's experiential learning theory has been criticised for its oversimplification (Ker and Bradley, 2011), its weaker theoretical evidence than other learning theories (Jarvis, 1987), and being grounded only in the individual learner and neglecting the social context of the experience (Bleakley, 2006; Yardley *et al.*, 2012). Further, it has also been questioned for the scarce information it provides about the types of experiences that may promote the students' engagement in the cycle (Morris and Blaney, 2011 p. 73). However, and despite reflection being a key part of experiential learning (Fry *et al.*, 2009), the need to examine and analyse experiences through the process of reflection, is insufficiently covered in Kolb's studies (Boud *et al.*, 1985; Jarvis, 1987). Thus for example, students decision making while performing or team learning "concrete experiences", such as those studied in the present research, would need a broader

underpinning than just a reflection *after* the event. Consequently, a learning theory that offers a deeper analysis of the reflection process will be discussed next.

Reflective practice learning

“Reflective practice” is a relatively new phrase that came into use particularly as a result of the work of Donald Schön (Moon, 2013 p. 80). The Schön theory of reflective practice (1983; 1987) is based on the reflection-change model that considers reflection as leading to action and then change. Further, he contends that reflection is a central element of professional thinking as this is how professionals deal with complex and ill-defined problems. So, the importance of Schön’s model lies in that dental and other healthcare practitioners are usually confronted with this kind of unclear clinical case.

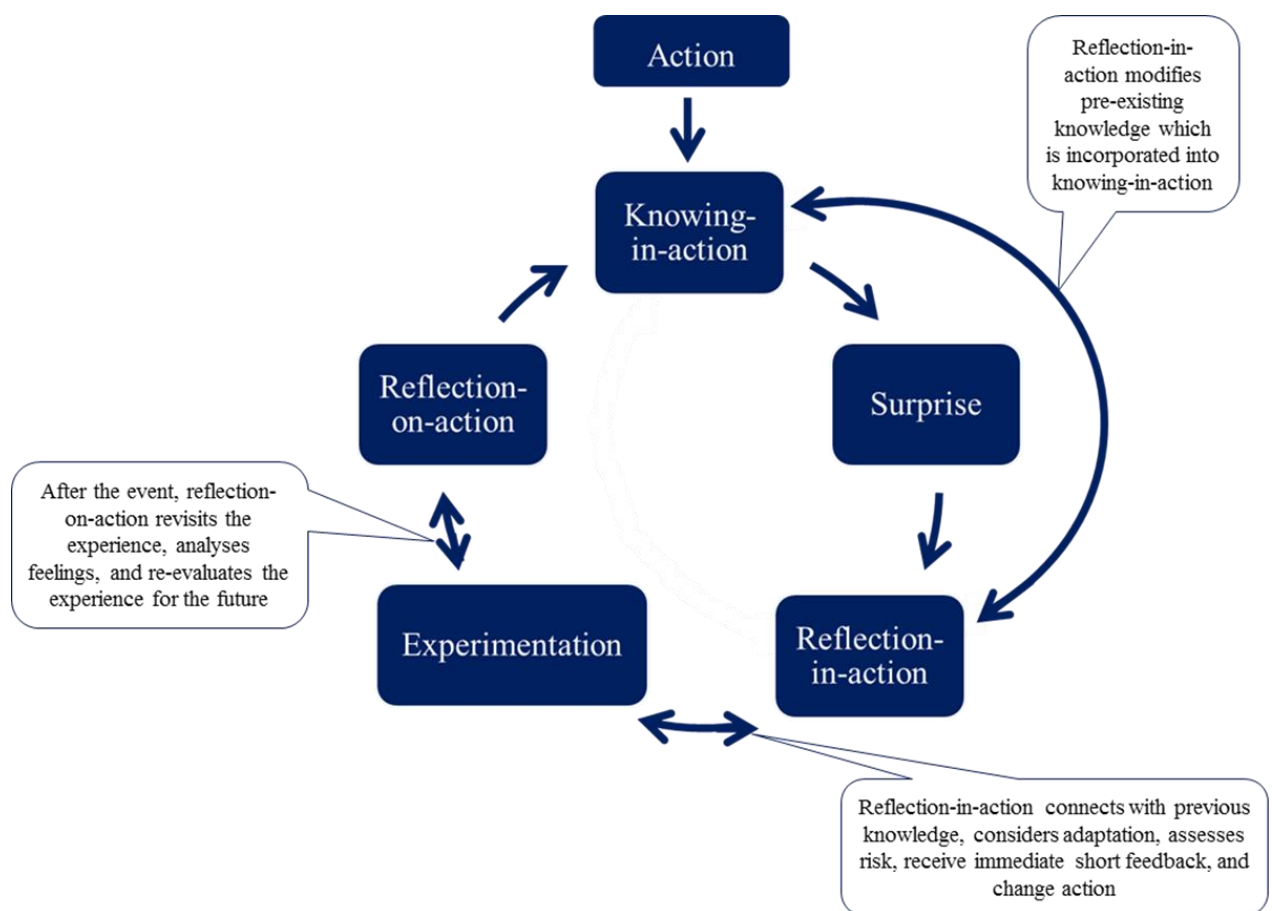


Figure 1-3 Five stages of Schön iterative model of reflective practice following a given action [interpreted from Schön (1983)].

In a different position to Kolb’s model (Kolb, 1984), where it is required to go through the complete learning cycle in order to further develop one’s own theory, Schön’s reflective practice model implies learning by simply reflecting critically during the

experience, ‘thinking on our feet’ (Schön, 1983 p. 54). In other words, there is no need to make a ‘mistake’ to reflect upon it. It is only required to readjust the theory through a double-loop learning (Figure 1-3). That is, reflecting upon the theory while in action, without going through the entire cycle of reflection after the event (single-loop) (Finger and Asún, 2001 p. 45).

Schön’s model has been described as highly relevant in healthcare education (Taylor and Hamdy, 2013), as it helps to explain what happens when reflective students interact with patients: they realise their limitations when confronted with challenging experiences; they improvise solutions; then later reflect on what happened; and finally consider how to respond in future similar experiences (Yardley *et al.*, 2012).

Thus, as Schön argues, experienced professionals base their practice largely upon ‘tacit knowledge’, that is, somewhat automatic responses grounded on existing mental schemas that enable them to perform efficiently in daily actions. This is termed professionals’ “knowing-in-action” (Schön, 1983), and has been paralleled to riding a bicycle (Shapiro and Talbot, 1991). Occasionally the bicycle skids, which could be the analogy to an unexpected problem or surprise of a professional in practice. At this time, a process of rethinking the “knowing-in-action” is triggered which makes us restructure the strategy, understand phenomena, or ways of framing problems. This gives rise to “on-the-spot experiment and further thinking that affects what we do in the situation at hand”, and probably later in similar ones as well (Shapiro and Talbot, 1991). This rethinking process is known as “reflection-in-action”, which Schön referred also as to ‘*professional artistry*’ (Schön, 1987 pp. 22, 28-29). Beyond names, the concept of “reflection-in-action” helps us to understand the decisions we take when faced with the above mentioned ill-defined clinical problems.

Further, Light *et al.* (2009 p. 14) assert that “reflection-in-action” is an ability to employ professional knowledge in complex events and situations, implying the need for knowledge, whilst for Knowles (2012 p. 188) it is the process of reflection while performing, used when existing mental schemas are no longer appropriate, that is, the checking and modifying of practice at the same time. Reflection-in-action requires high cognitive awareness to decision-making and problem-solving in parallel, and as such is less likely to be biased by emotions and subjectivity compared to the automated responses of knowing-in-action (Barley, 2012).

In contrast to experts, novices practice using conscious thoughts and decision making

(Barley, 2012), as they lack professional “knowing-in-action” (they are learning to ride the bicycle). Thus, they tend to cling to rules and procedures, which they normally apply in a mechanical way (Finlay, 2008). However, if students base their practice in the automatic responses of knowing-in-action, they will miss opportunities to think about what they are doing (Mamede and Schmidt, 2004).

Consequently, while teaching reflective learning to novices, it is important to instil in them the need to think while practicing and solving problems (thinking on your feet). Thus, by establishing a reciprocal feedback dialogue of reflection-in-action, they learn from the feedback, change and therefore learn effectively (Schön, 1987; Knowles *et al.*, 2012).

However, Schön also proposed the construct of reflection-on-action which occurs later on, after the event (Schön, 1983 p. 278), and provides opportunities to learn from the earlier decision making process (Mamede and Schmidt, 2004). This is a method that both experts and novices could adopt as a reconstructive and conscious mental review, analysis and evaluation of past experiences to determine what may have contributed to the unexpected, and how this situation may affect future similar events (Kaufman and Mann, 2010).

Both reflection-in-action and reflection-on-action are recurring processes that modify pre-existing knowledge directly (Barley, 2012), and where insights and learning from experiences may be incorporated into future “knowing-in-action” (Schön, 1983; Schön, 1987) (Figure 1-3).

Schön’s theory of reflective practice has been criticised for the lack of conceptual clarity surrounding the term “reflective practice” (Ruth, 2013), as it may refer to “a complex array of cognitively and philosophically distinct methods and attitudes” (van Manen, 1995). Further, the emphasis on individual reflection (Sandywell, 2013 pp. 249, 253) and the questionable evidence to distinguish between reflection in- and on-action (Eraut, 1995), have also created disagreement. However, as the very same author asserts, Schön’s book “The Reflective Practitioner” (1983), is the most quoted book on professional expertise.

Andragogy

The term andragogy was introduced by the adult educator Malcolm Knowles as “the art and science of helping adults (or, even better, maturing human beings) learn” to

differentiate adult's learning – andragogy – from children's learning - pedagogy - (Knowles, 1980), as he considered they learnt differently. So, andragogy emphasises that adults are self-directed individuals and as such expect to take responsibility for their decisions (Merriam *et al.*, 2007).

Knowles presented his model as a set of four 'assumptions' (a fifth and a sixth were later included) (Knowles *et al.*, 2012 pp. 138-140), and not as an empirically based theory (Kaufman and Mann, 2010). Despite this, Knowles' model has produced big controversies. While Norman (1999) argues that "adult learning" (probably avoiding the term 'andragogy') is not a theory, Merriam (2001 p. 8) highlights the impact and usefulness of andragogy "for exploring some of the definitional and philosophical issues related to the evolution of adult education as a scientific discipline, and its strengths and weaknesses as a guide to practice". Again against it, Hartree (1984 p. 203) stated that its popularity can be attributed to "the fact that it makes the kinds of points that adult educators tend to support instinctively for emotional reasons", that is, Knowles "says what his audience wants to hear".

Andragogy as a model is related to several educational, social, philosophical and psychological theories that were clustered by Knowles to clarify that adults learn differently and have particular attitudes towards learning (Taylor and Hamdy, 2013). Accordingly, Knowles *et al.* (2012 p. 140) clearly identify andragogy as being rooted in humanistic perspectives, primarily concerned with the self-actualisation of the individual, and pragmatic philosophy, where knowledge from experience is valued over the one from formal authority. Further, in andragogy, feedback is usually more important than tests and evaluations (Steinert, 2011), and should be delivered in small doses, with the opportunity to obtain more if needed, in order to stimulate the self-assessment process of learning needs (Knowles *et al.*, 2012 pp. 124, 322).

Probably due to an early criticism, enormous debate stating that his pedagogy and andragogy differentiation was artificial (Taylor and Hamdy, 2013), Knowles later modified his model by describing andragogy and pedagogy as a learning continuum (1984). However, his model has played an important part in freeing adult learning from being treated like children in an appropriate way (Yardley *et al.*, 2012).

The changing emphasis, problems and strategies at different times, that underline both approaches as described by Knowles *et al.* (2012) are summarised in Table 1-1.

Table 1-1 Pedagogical and Andragogical models comparison according to Knowles *et al.* (2012 pp. 298-299).

Pedagogical Model	Andragogical Model
The learner is a dependent personality. The teacher/trainer decides what is to be learnt, how and when it should be learnt, and whether it has been learnt.	The learner is self-directing. Adult learners want to take responsibility for their own lives, including planning, implementing and evaluating of their learning activities.
The learner has little experience that can be used in the learning process. The experience of the teacher/trainer is what is important. Thus, one-way communication strategies are employed to transmit information.	The learner enters an educational situation with a great deal of experience, which is a valuable resource to the learner as well as to others.
People are ready to learn when they are told what they have to learn in order to advance to the next grade level.	Adults are ready to learn when they perceive a need to know or do something in order to perform more effectively in some aspect of their lives. They can be stimulated by helping them to assess the gaps between where they are now and where they want to and need to be.
People are motivated to learn primarily by external pressures from parents, teachers/trainers, employers, the consequences of failure, grades, certificates, etc.	Adults are motivated to learn by internal factors after they experience a need in their life situation. Learning needs to be problem-focused or task-centred, to apply the new knowledge as quickly as possible.

Beyond the controversy, andragogy principles can be incorporated into the design of any educational programme to enhance the learner's receptivity, relevance and engagement (Merriam *et al.*, 2007). Further, an understanding of the assumptions can also influence pacing, meaning and motivation (Steinert, 2011).

The described learning theories (Figure 1-4) have specific characteristics that further define how individuals learn, and have implications for all of us who have a role as teachers/tutors to become professional educators and thus facilitate and maximise students' learning.

These theories can provide a guiding framework to understand a given teaching approach and better implement an academic programme. Some teachers teach how they have been taught, while others acquire tacit knowledge of pedagogical principles, perhaps by reflecting on their own teaching events. Fortunately, as posited by McLeod *et al.* (2008), it is today well recognised that content expertise is not enough to succeed as an educator. Thus, it is accepted that understanding the learning process by knowing educational theory, will result in an improved teaching performance (McLeod *et al.*, 2008; Foster and Laurent, 2013).

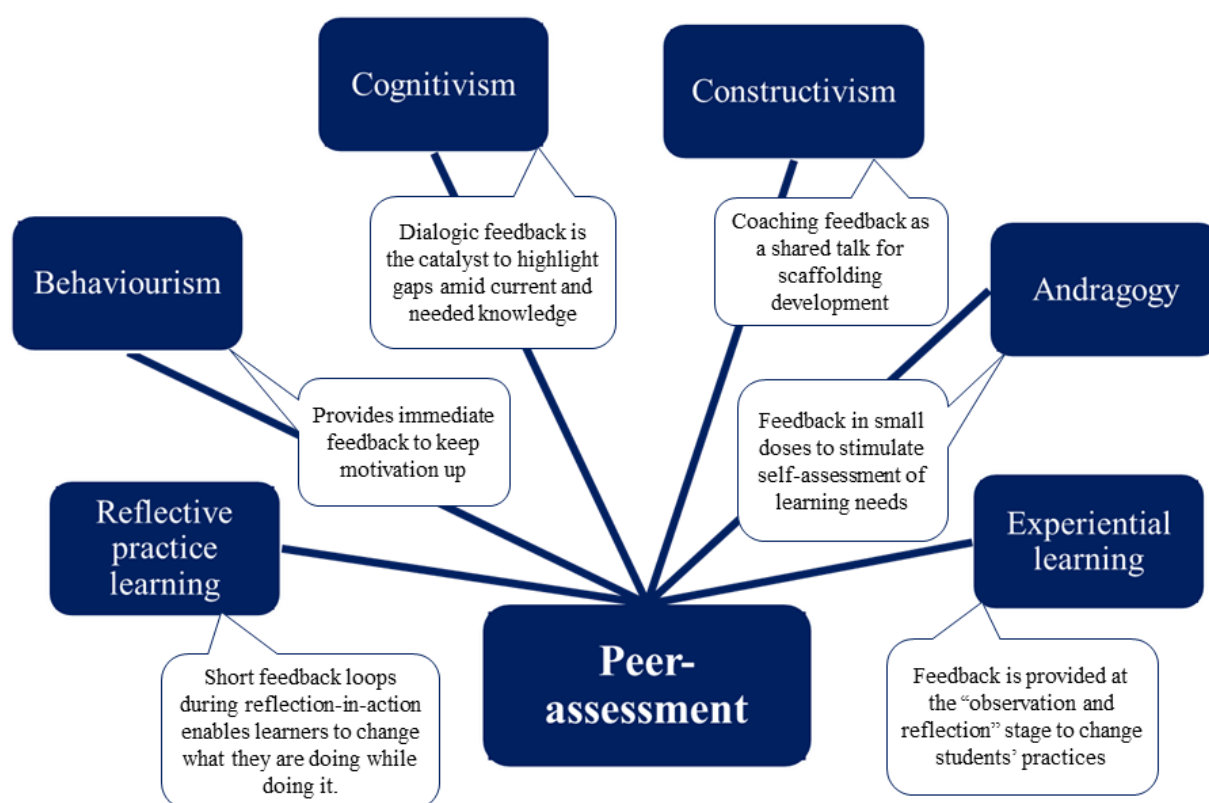


Figure 1-4 Graphical interpretation of learning theories related to peer-assessment through feedback.

1.3 Learning Theories and the Curriculum

Learning theories have a reciprocal relationship with curriculum models and their practical effects on educational practice are very powerful (Grant, 2010). The model and structure of the curriculum is a fundamental element of every educational institution and dental schools are no exception (Oliver *et al.*, 2008).

A simple definition for curriculum, as presented by Kern *et al.* (2009), is “a planned learning experience”. Somewhat more elaborated is the one from The Institute for International Medical Education that defines curriculum as “an educational plan that spells out which goals and objectives should be achieved, which topics should be covered and which methods are to be used for learning, teaching and evaluation” (Wojtczak, 2002). Moving away from a static concept, in the 1970s Stenhouse (1975) stated that the curriculum should be “open to critical scrutiny”. This allows the interpretation of the curriculum as a dynamic educational plan as ideas change. In fact, both learning theory and pedagogical practice, together, are seen today as a never-ending work in progress (Grant, 2010).

In the 1960s a medical curriculum was expected to contain objectives and experiences based on the needs of the society, the philosophy of scientific thinking and the professional characteristics of physicians (Miller, 1961). Later in the 1970s, when behaviourism was predominant, the advice was to define the curriculum in terms of objectives in behavioural terms shaped by the teacher (Simpson, 1972). Subsequently, the 1980s witnessed a move away from the teacher as the transmitter of knowledge, characteristic of cognitivism (Knowles *et al.*, 2012), towards a systems approach (context bound) that emphasised teaching methods with an active participation of students (Newble and Cannon, 2001), more typical of a constructivist perspective (Biggs and Tang, 2011). Attention has moved during the last 30 years from ‘teaching’ towards ‘learning’ (Grant, 2010).

This brings us back to the last part of Wojtczak’s (2002) definition of curriculum. That is, the ‘learning’ and ‘evaluation’ (assessment). As will be discussed below, the focus on learning is today closely related to assessment, and more specifically to formative assessment or ‘for learning’. However, this is not always the case. For example, in the behaviourist model of curriculum, where the teacher is the source of knowledge, assessment takes the role of a ‘test’ at the end of the process for students to show what they have learnt, that is ‘assessment of learning’ (Fish and Coles, 2005). By contrast, in curriculum models where students take an active role in their learning process and teachers turn out to act more as facilitators, as in a constructivist approach, assessment becomes an ongoing process of understanding of the student achievements, informing their next educational needs. Thus, in this model, assessment for learning is a critical part of the teaching and learning process (Fish and Coles, 2005).

The fact that learning is the central purpose in these curricular models, the development of modules or courses can adopt a more student-focused approach. This would mean that the teacher designing a module would start thinking about what the students need to learn, rather than what she or he will be teaching. The key principle is to start by designing the assessment of the module followed by the design of the content of that module (Fry *et al.*, 2009). Further, the assessment should be part of an integrated *system* of assessment (Schuwirth and van der Vleuten, 2010), which, as described below, would serve both “assessment for learning” and “assessment of learning” purposes (Ramsden, 2003). The evidence suggests that student-centred models, where the purpose of teaching is to support learning, motivate the learner to adopt a deeper approach to learning (Biggs and Tang, 2011). This model has been described as a useful

and efficient approach to curriculum design in dentistry (Manogue and Brown, 2007).

Despite the strong influence learning theories have on curriculum design, there are a wide variety of factors affecting curriculum planning, for example, resources, culture and ethnographics, available learning environments, healthcare service, regulations, and society as a whole (Oliver *et al.*, 2008; Grant, 2010).

1.4 Assessment

Assessment is generally acknowledged as an essential and fundamental part of the education process (Schuwirth and van der Vleuten, 2010), and it is not only important for the student but also for the teacher and course organiser, the accrediting body and the public as a consumer (Harden and Laidlaw, 2012). Assessment provides a window into what students know and ignore, and how they are thinking (Earl, 2012). It affects their lives as their future directions and careers depend on it (Boud and Falchikov, 2007).

The assessment of students learning has been seen as a work overload both to students and tutors, a process that hinders learning, encourages superficiality and conformity, and that it only needs to be done for universities to maintain standards (Knight, 2012). This despite the evidence that started to be available almost twenty years ago (Barr and Tagg, 1995), in that “students’ assessment is at the heart of an integrated approach to student learning”. Further, by then the shift from universities as institutions to “provide instruction” to one that “produce learning”, was already taking hold (Barr and Tagg, 1995). Today, it is widely accepted that assessment, rather than teaching, has a major influence on students’ learning (Boud and Falchikov, 2007). However, the debate about assessment continues and has now moved to issues such as academic standards, preparing students for employment, measuring quality and providing incentives (Boud and Falchikov, 2007).

This is also reflected in the many different roles “assessment” can take (Harlen, 2007), ranging from a certification procedure leading to a pass/fail decision, to assessment as an evaluative or feedback action in education (Manogue *et al.*, 2002; Boud and Falchikov, 2007; Harlen, 2007). As proposed by Schuwirth & van der Vleuten (2010), assessment in competence-based education is “any purported and formal action to obtain information about the competence and performance of a candidate”. In a like

manner, for Fish and Coles (2005 p. 169) assessment is an “educational activity of recognising and recording learners’ achievements and their development within a specific context and in the light of the quality and scope of the education provided for them”.

Thus, assessment is always undertaken with a specific purpose which unfortunately most see just as producing a mark (Moon, 2013), forgetting it also has four main roles: pedagogy (promote learning), measurement (evaluating competence), standardisation (marking to make progress decisions), and certification (fitness to practice) (Fry *et al.*, 2009 p. 134). Moreover, it is often perceived as an afterthought and a burden at the end of the teaching process. Fortunately, a gradual shift is being observed with an increasing attempt to teach what we assess and assess what we want to train for (Crossley and Jolly, 2012). The rationale for this is the widely accepted notion that “assessment drives learning” (Miller, 1990; Wass *et al.*, 2001; Schoonheim-Klein *et al.*, 2006; Manogue *et al.*, 2011; Norcini *et al.*, 2011; Dolmans and Tigelaar, 2012; van der Vleuten *et al.*, 2012). Thus, if teaching and assessment contents comprise an authentic representation and are within the boundaries of the defined competences and learning outcomes (Biggs and Tang, 2011; Tavakol and Dennick, 2011), they will be sending the students the right signal for them to know where, why and what they are working and learning for, directing their learning towards those desirable outcomes (Oliver *et al.*, 2008; Biggs and Tang, 2011; Manogue *et al.*, 2011). Consequently, assessment is a key and integral part of the curriculum and should be seen as inseparably linked to the learning outcomes and teaching methods (Harden and Laidlaw, 2012). As Melnick (2002) put it, “the curriculum instructs teachers; the exam instructs students what to learn”.

1.4.1 Assessment purposes

Assessment can either be summative, to measure students’ achievements, or formative, to enhance their learning (Light *et al.*, 2009). However, some scholars argue that this distinction has become blurred as summative and formative purposes can be mixed in one single assessment (Schuwirth and van der Vleuten, 2010; Harden and Laidlaw, 2012). Others, however, have called for caution, as if the same assessment is used for both formative and summative purposes, it creates a conflicting situation for the students: “they are being asked to display and to hide error simultaneously” (Biggs and Tang, 2011 p. 197). However, and despite agreeing with this latter statement, it does not mean students cannot get “formative” feedback from a “summative” assessment.

Summative assessment is often referred to as assessment *of* learning and is used to confirm what students know for evidence based decision making or certification purposes (e.g. criteria-referenced assessment) (Sadler, 2005). For example, it can be used to rank students, to select who is admitted to a given programme (norm-referenced admission tests), to make judgments about students' competence progression at a defined level within their programme, or to certify whether students' are 'fit for purpose' when they complete the programme and have achieved the regulating bodies expected standards (Light *et al.*, 2009; Harden and Laidlaw, 2012). Assessment *of* learning becomes public and it affects students' futures. Therefore, it is important that these assessment measurements are credible and defensible.

Historically, educational assessment has largely been assessment *of* learning (Earl, 2012), as it is an essential part of education (Harlen, 2007), and despite some intense criticism, mainly on its impact on students' learning, the need for certification has secured the "high-stakes" (Knight, 2006) summative assessment position (Boud and Falchikov, 2007 p. 4). Caution has been highlighted for exam-dominated systems, as students' strategy to pass examinations will become more important than gaining knowledge, leading inevitably to surface learning (Biggs and Tang, 2011).

However, a counter-movement which probably started with the review of classroom formative assessment by Black and William (1998), has seen an emphasises on assessment *for* learning and how it influences students' learning (Boud and Falchikov, 2007). Further, contemporary educational philosophies state that assessment *for* learning should take priority over assessment *of* learning (Schuwirth and van der Vleuten, 2011b). Indeed, the word 'assessment', is derived from the Latin *ad sedere* or *assidere*, "to sit beside or with", and emphasises the importance of the feedback component of assessment (Wiggins, 1993; Manogue *et al.*, 2011).

Formative assessment involves students' development, improvement and learning (Rolfe and McPherson, 1995; Light *et al.*, 2009), and it is used to find out the level of students' understanding to provide them with descriptive and informed feedback on their performance progress, to motivate and guide their future learning (Epstein, 2007; McDowell *et al.*, 2010; Manogue *et al.*, 2011; Harden and Laidlaw, 2012).

However and despite students receiving feedback, it is often the case that they do not understand what it means, and thus, cannot take any action until they share the feedback provider conception (Boud and Falchikov, 2007; Bloxham, 2009; Boud and Molloy,

2013).

Consequently, recent research on feedback as part of a formative assessment has gone deeper focusing on the use the student gives to the provided feedback. This, the “ultimate goal”, has been coined assessment *as learning* (Earl, 2012). Accordingly, in addition to receiving external feedback, assessment *as learning* emphasises this, the feedback, as a process of metacognition for students to critically self-assess and provide their own feedback so that they become their own best assessors fostering their self-regulation process (Earl, 2012). In assessment *as learning*, feedback is not just a transfer of ideas or a judgement where and what to improve (Bloxham, 2009), it requires the active process of critical self-assessment so that students create their own understanding, making sense and relating these new transferred ideas to previous knowledge, for an action (Sadler, 1989) and effective learning to occur (Black and Wiliam, 1998). To be successful, this process requires the students to be aware of the standards they are aiming for (Earl, 2012), so that they can judge their own performance in relation to those standards. Involving and giving them some responsibility in the assessment process provides them with the opportunity to learn those standards in a given context (Bloxham, 2009). These concepts, in that students should not just be passive receptors of feedback, have also motivated some scholars to call for a re-examination of feedback in higher education (Nicol, 2010; Sadler, 2010). Further, the concept of continuing development at work (once graduated) requires individuals to be creative, to seek and utilise feedback to increase their productivity and be effective practitioners, and to be able to this, universities need to prepare them with high levels of self-regulatory ability (Boud and Molloy, 2013).

Given its importance to the current study, feedback is discussed separately under its own heading below (page 62).

1.4.2 Assessment principles

The design and setting up of a successful assessment system is not easy. There are many different assessment instruments (each with advantages and disadvantages), described in the vast published research literature on medical/dental education (Manogue *et al.*, 2002; Epstein, 2007; Albino *et al.*, 2008; Kramer *et al.*, 2009; Manogue *et al.*, 2011; Norcini *et al.*, 2011). Teachers should understand the outcomes to be assessed and the need for a blend of assessment methodologies (Manogue *et al.*, 2011) as that “perfect assessment is an illusion” (van der Vleuten, 1996). That is, no single methodology can

test all competencies and performances of Miller's pyramid layers (Figure 1-5) (Miller, 1990; Epstein and Hundert, 2002; Carr, 2006; Chadwick and Holsgrove, 2009; Davies *et al.*, 2009; Schuwirth and van der Vleuten, 2010).

In order to overcome each assessment method weaknesses (Schuwirth and van der Vleuten, 2010), together with psychometric studies highlighting assessors' subjectivity and students' case-specificity of clinical performance (Crossley *et al.*, 2002b), during recent years there has been a move towards multiple 'mini' test samples across many different assessment formats, such as the Objective Structured Clinical Examination (OSCE) and a wide variety of Workplace-Based Assessment (WPBA) tools (Norcini *et al.*, 1995; Norcini *et al.*, 2003; Prescott-Clements *et al.*, 2008; Schuwirth and van der Vleuten, 2011a; Crossley and Jolly, 2012).

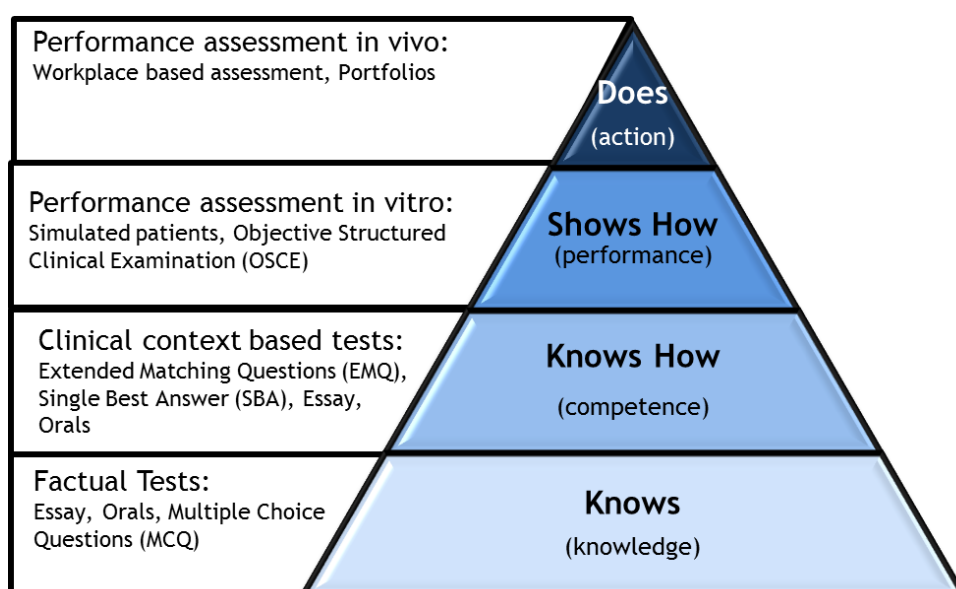


Figure 1-5 Miller's pyramid for clinical assessment (Miller, 1990) and frequently used assessment methodologies (Kramer *et al.*, 2009).

However, regulating authorities of education and training programmes demand different strengths of assessment methods (General Dental Council, 2012c). Thus, the widely accepted criteria used to evaluate the strengths of a given assessment method (Watson *et al.*, 2014) proposed by van der Vleuten (1996), includes its reliability, validity, cost and feasibility, acceptability, and educational impact. As the peer-assessment protocol developed and implemented in the subsequent research chapters will be judged against these principles, they will be described in the following pages.

Reliability

An assessment method is reliable when its results are reproducible, that is, the extent to which a test yields the same results on repeated trials (Carmines and Zeller, 1979; Schuwirth and van der Vleuten, 2010), and thus its value will reflect the amount of error, both random and systematic, which are intrinsic to any measurement (Streiner and Norman, 2008). This sampling ‘error’ appraisal is particularly important in assessment methods where candidates are assessed by more than one assessor, for example in an OSCE or longitudinal clinical assessments. Thus, the intra-rater and the inter-rater reliability coefficient will estimate the consistency of ratings within the same observer and between them, respectively (Streiner and Norman, 2008). Psychometric studies have concluded that in these cases, multiple examiners across different cases improve inter-rater reliability (Wass *et al.*, 2001).

According to assessment theory, one of the easiest ways to estimate the reliability of, let’s say, a written assessment, is the test-retest analysis in which the same test is given to the same students in two separate occasions. The correlation of both scores on the two administrations will provide the reliability of the test. If the results are exactly the same, the reliability coefficient will be 1.00. However, invariably, the correlation will be less than perfect (Carmines and Zeller, 1979). Thus, reliability coefficients of ≥ 0.7 and ≥ 0.8 are generally accepted for low and high-stakes situations, respectively (Beard *et al.*, 2011).

Another option to estimate reliability of an assessment is the parallel test in which, for example, two groups of candidates are assessed using different questions of the same field that are thought to be equivalent in difficulty (Wass *et al.*, 2001). As before, the correlation between both scores will provide the reliability of the test (Schuwirth and van der Vleuten, 2010).

However, both test-retest and the parallel test are not practical. Consequently, most of the well-known reliability tests in Classical Test Theory (CTT), like Cronbach’s alpha (Cronbach, 1951), use a retrospective approach in which the test is randomly divided into two halves and treated as parallel tests. Thus, in CTT the main assumption is that a student has a true ability (known as “true score”), but due to measurement error (known as “error score”), which will always be present, the candidate does not obtain the exact same score even if the same test is used twice. Consequently, the “obtained score = true score \pm error score (DeVon *et al.*, 2007). Then, as explained before, the similarities

between these scores are used to mathematically calculate the reliability. The higher the similarity, the lower the “measurement error” (Streiner and Norman, 2008). Further, CTT works best with multiple-choice questions (MCQ) tests since all students take the same questions, as it focuses on the test and its errors (Schuwirth and van der Vleuten, 2010). However, in the presence of several errors influencing the results of a test, like multiple assessor and students being assessed with, for example, different patients or parallel OSCE’s lines, another reliability theory, namely Generalizability Theory (GT), is said to work better (Tavakol and Dennick, 2012).

GT is therefore used when there are variabilities due to multiple examiners (e.g. hawks and doves), clinical scenarios (e.g. different patients or OSCE stations), and the known students’ clinical case-specificity (Eva *et al.*, 1998), where some do better in some cases than their classmates. Generalizability Theory includes all such components and provides an equivalent to reliability called Generalizability Coefficient (Schuwirth and van der Vleuten, 2010). Through sophisticated calculations, GT allows knowing, for example, whether we need more stations or fewer examiners, to reach a certain level of Generalizability. However, and despite a high number of items in a given test taken by a large numbers of students will get better G Coefficient, they cannot be easily extrapolated to a new situation (Schuwirth and van der Vleuten, 2011a).

The third theory concerned with the measurements of reliability is Item-Response Theory (IRT). As opposed to CTT and GT, IRT can be used to identify the behaviour of individual test items and how they interact with individual student abilities (Tavakol and Dennick, 2012), that is, estimate item difficulty independently of students’ ability, and vice versa. Thus, IRT, which is best used with one thousand or more students, can determine if a low score in a given test is the result of particular items of the test being too difficult or the students’ low ability (Tavakol and Dennick, 2013). Further, the reliability of a test consisting of previously used items from an item bank, or students of different abilities, can be calculated before the test is delivered (Schuwirth and van der Vleuten, 2010).

Validity

In a general sense, any measuring device is ‘valid’ if it does what it is intended to do, and does it cleanly without accidentally including other factors (Carmines and Zeller, 1979). In educational terms, validity is the extent to which the competence that the assessment claims to measure is actually being measured (Schuwirth and van der

Vleuten, 2010). This requires the collection of evidence from different approaches, and the focus is not necessarily on scores or items, but rather inferences made from the instrument (Gregory, 2004 pp. 97-98) Therefore, these inferences can be classified and better understood examining some facets of validity that will be studied later in the research chapters. These are content validity, construct validity, and criterion validity (Streiner and Norman, 2008).

Content validity refers to whether the items or questions on a given test accurately reflect the whole testable domain, and therefore it is not determined by its format but by its content (Schuwirth and van der Vleuten, 2010; Schuwirth and van der Vleuten, 2011a). In other words, are items in the test an effective representation of all possible questions that could be derived from the content? (Gregory, 2004), and there are no irrelevant items (Streiner and Norman, 2008 p. 250). For example, an examination on ‘cariology’, would not have content validity if it only comprises questions on, let’s say, “affected dentine”, neglecting “infected dentine” and other ‘contents’. To ensure adequate and relevant coverage of a given test, a matrix, called ‘blueprint’ (Crossley *et al.*, 2002b) is frequently used to efficiently identify the domain(s) of interest, the appropriate assessment methods, and the number of items per domain or category (Schuwirth and van der Vleuten, 2010). However, despite the importance of content validity, it is not sufficient to ensure ‘validity’ of the test, as it does not provide any evidence from the test scores (Messick, 1995; Streiner and Norman, 2008).

To address this issue, the concept of “construct validity” was introduced by Cronbach and Meehl (1955) almost 60 years ago. Construct validity is a back-to-front approach to hypothesis testing (Crossley *et al.*, 2002b). For example, fourth-year dental students’ manual dexterity, which would be the ‘construct’, might be expected to be better than that from first-year dental students, which would be the ‘hypothesis’. Accordingly, a test measuring students manual dexterity applied to both cohorts, would demonstrate “construct validity” if it shows that fourth-year dental students perform better than first-year ones.

Consequently, construct validity indicates the extent to which the scores from the test support the devised hypothesis based on the knowledge of the underlying construct (Norman and Eva, 2010). Subsequently, every time a test is implemented in different situations, let’s say the manual dexterity test in postgraduate trainees, new hypotheses could be derived, and so construct validity would be a continual task (Streiner and

Norman, 2008).

Further on validity testing, criterion validity is the correlation of a test result with some other measure of the attribute being studied, ideally a “gold standard” which is accepted in the field (Streiner and Norman, 2008 p. 254). Accordingly, a high degree of correlation between the “gold standard”, or criterion variable, and the scores on the testing instrument, would be evidence of criterion validity (DeVon *et al.*, 2007). This is sometimes referred to as the “validity coefficient” (Carmines and Zeller, 1979 p. 18; Gregory, 2004 p. 101).

Cost and Feasibility

Good assessment is definitely costly, mainly due to staff training and their time spent on planning, taking and marking the assessment (Boursicot *et al.*, 2011). This is especially true “if judgements are any finer grained than pass-fail” (Knight, 2007 p. 77). Unfortunately, economies of scale are difficult to achieve as most assessment costs are directly proportional to student numbers (Gibbs and Simpson, 2004).

However, it should be borne in mind that investing in assessment is investing in teaching and learning (van der Vleuten, 1996), and that expensive assessment can still be good value (Knight, 2007), so far assessment costs do not overtake teaching costs.

Studies reporting assessment methods feasibility in healthcare education do this with different criteria. Accordingly, the systematic review by Donnon *et al.* (2014) found that within the heterogeneity of Multi-Source Feedback (MSF) instruments evaluated, feasibility was primarily based on the response rate percentages, and rarely included costs and administration concerns. Similarly, other studies have estimated the feasibility of assessment methods by means of completion rate, completion time, faculty time required for review, to score, and to provide feedback, and satisfaction rating (Ram *et al.*, 1999; Kogan *et al.*, 2003; Torre *et al.*, 2007; Mori *et al.*, 2008; Durning *et al.*, 2012; Tolsgaard *et al.*, 2013).

Acceptability

Acceptability is another quality required for sound assessment practice (Hays *et al.*, 2002). Even the best assessment method is useless, and it will not survive (van der Vleuten, 1996), if it is not accepted and credible by both teachers and students (Schuwirth and van der Vleuten, 2010; Norcini *et al.*, 2011).

Further and specifically for students, an assessment method will be better accepted when they perceive its educational value and fairness (Norman *et al.*, 1991). This was the case in the study by Escudier *et al.* (2014) where a big step was undertaken to control cheating in high-stakes e-assessment of dental students (by introducing privacy screen filters), which for 86% (119/138) of students it increased test fairness. Conversely, Tonni and Oliver (2013a) piloted a reflective e-portfolio in six orthodontic postgraduate trainees who highlighted acceptability issues related to it. In line with the literature, the authors acknowledged that the future successful implementation of the new instrument will depend upon the solution of the acceptability issues identified by both students and mentors.

Educational impact

As discussed in the previous section (Assessment purposes in page 43), the educational impact of an assessment method or programme is exemplified by the saying “students don’t do what you *expect*, students do what you *inspect*” (Schuwirth and van der Vleuten, 2010 p. 198).

Therefore the driving force of assessment should be exploited to achieve the educational objectives (van der Vleuten, 1996). This must consider the principle of “constructive alignment” between the educational objectives and the assessment objectives (Biggs and Tang, 2011). When they are not aligned, the assessment objectives will prevail (van der Vleuten, 1996).

The degree of educational impact of a teaching intervention is increasingly being expressed in the research literature as the “effect size” (Hojat and Xu, 2004). The effect size, known as d (average post-test - average pre-test / spread) (Hattie, 2012 p. 271), has been described as a useful method for comparing the mean results on different measures, or over time, or between groups (Hattie, 2012), independently of the study sample size. The effect size has become increasingly popular as it can be used as a supplement or an alternative to the statistical significance analysis (Fan, 2001).

The average effect size of schooling on overall student achievement, considering more than 100 factors (including student, home, school, teacher, curricula, and teaching domains), has been reported by Hattie and Timperley (2007) to be $d=0.40$. This means that schooling increases the mean on an achievement test by 0.4 of a standard deviation. Further, this value is today used as a “gold standard” on which to judge the effects of

any educational study (Norcini and Burch, 2007). In a later publication by Hattie based on more than 900 meta-analyses of +50.000 research articles with about 240 million students, he suggests that an effect size of $d < 0.20$ would be small, between 0.30 and 0.60 would be medium, and > 0.60 would be large (Hattie, 2012 p. 13).

It should be noted that Cohen (1988 pp. 25-27) had previously classified effect size values in the same three categories but with some differences in their values. For Cohen, a small effect size would be $d=0.20$ (negligible practical importance); medium would be $d=0.50$ (moderate practical importance); and large would be $d=0.80$ (crucial practical importance).

1.4.3 Is there a role for peers in assessment?

The *a priori* answer to the above question is yes, provided it is understood in terms of a dialogue, rather than just a substitute for tutor assessment and feedback (Hamer *et al.*, 2014; Nicol *et al.*, 2014). Further on the rationale for this positive answer, it is today well documented that peer-assessment followed by peer-feedback helps students to learn more effectively (Topping, 2005), as the peer experience should not only provide insights into what they know and can do (limiting the encounter to just scoring), but also steer them to improve (Friedlander and Anderson, 2011). Students not only gain a deeper insight into subject matters, but a significant “ability to engage with and take ownership of evaluation criteria, to make informed judgements about the quality of the work of others, to formulate and articulate these judgments in written form and, fundamentally, the ability to evaluate and improve one’s own work based on these processes” (Nicol *et al.*, 2014).

Peer-assessment can be defined in an educational framework as an arrangement that involves observation of students who have attained the same general level of training or expertise and status in order to judge structured tasks or provide global impressions of the amount, level, value, worth, quality or success of their peers’ work (Topping, 1998; Norcini, 2003a; Finn and Garner, 2011). As a result, students are required to provide their peers with grades, feedback or both (Boud and Falchikov, 2007). Indeed, it has been reported that formative peer-assessment can successfully result in the provision of objective feedback (Falchikov and Goldfinch, 2000; Sargeant *et al.*, 2011), which can benefit and enhance the students’ learning process in several ways.

As discussed below, peer involvement in assessment during higher education can aid

learning when it takes place, but it can help to prepare students for independent and autonomous work as well, and by doing so, for a lifetime of learning (Boud and Falchikov, 2007). Further, peers are a key feature of learning in the workplace and in professional practice, as learning with and from peers is the dominant mode of everyday learning (Falchikov, 1995). Furthermore, peer-assessment has the potential to generate profits in the cognitive, social, affective, and professional skills domains (Topping, 1998), and in healthcare it has a long history of identifying successful practitioners (McCormack *et al.*, 2007).

In healthcare education, peers are in an advantageous position to judge one another's clinical competencies (McCormack *et al.*, 2007), as they have a frequent and close contact in a variety of contexts not always available to faculty (Al Khalifa *et al.*, 2013). This increased exposure (McDonald, 2010) while performing the tasks and procedures that are being learned under real conditions (Shumway and Harden, 2003), allows them to observe, assess and provide feedback (Sargeant *et al.*, 2011) to each other in a less stressful approach (Evans *et al.*, 2007). Furthermore, producing quality feedback as a fundamental graduate skill (Nicol *et al.*, 2014), is an area of practice that is important to develop in university education as students report feeling ill prepared when entering the medical workforce (Burgess *et al.*, 2013).

These potential advantages might explain the reason why peer-assessment has been recommended and encouraged in many publications (Manogue *et al.*, 2002; van der Vleuten and Schuwirth, 2005; Plasschaert *et al.*, 2007; Mattheos *et al.*, 2008; Kramer *et al.*, 2009; Finn and Garner, 2011; Manogue *et al.*, 2011; Tonni and Oliver, 2013b).

In summery, peer-assessment can be beneficial to promote learning if it (Falchikov, 2007 p. 139):

- Is designed as a learning experience.
- Requires learners to take responsibility for their actions.
- Involves providing, seeking and utilising feedback.
- Encourages a reflective approach to learning.
- Requires students to identify and apply standards and criteria.
- Provides some degree of modelling and/or scaffolding.
- Is practised in a variety of contexts.

1.4.4 Reported benefits of peer-assessment

Dochy *et al.* (1999) reviewed 63 studies and suggested that the use of a combination of new assessment forms such as peer-assessment to inform self-assessment, encourages students to become more responsible and reflective, as the process requires the students to be fair and accurate with the judgments they make regarding their peers. The enhancement of students' learning by means of reflection was likewise reported by Falchikov (1995). She also found that peer-feedback improved learning by encouraging analysis and diplomatic criticism, and again in agreement with Dochy *et al.* (1999), she later argued that peer-assessment was a necessary step to develop self-assessment skills (Falchikov, 2007).

The meta-analysis on peer-assessment in higher education conducted by Falchikov and Goldfinch (2000), found that peer-assessment of academic products and processes corresponded more closely to tutor ratings than did professional practice, probably due to students familiarity with the former skills. They recommended conducting peer-assessment in academic settings where students are normally involved, suggesting that peer-assessment could be successful in any discipline area and at any level.

A large analytical review of the literature on UK authorship of innovative assessment methods (from 1996 onwards) across the disciplines, was carried out by Hounsell *et al.* (2007). Among the more than 400 analysed studies, a number (N=58) were included in the "student involvement in assessment" category and the majority were peer-assessment. The most recurrent advice was the need for preparation and training; making clear the rationale for involving students in assessment; and using pilot studies. Additionally, implementation benefits included aiding the development of personal and lifelong learning skills, helping students overcome unrealistic expectations, facilitating shared understanding and encouraging integration. Likewise, Nulty (2010) found that peer- and self-assessment facilitate greater student involvement in their learning development, and called for a greater use of both methods in the first year of higher education and onwards.

According to Nicol and Macfarlane-Dick (2006), students who have just learned something are frequently better able to explain it to their peers than tutors, as they use a comprehensible language and a more friendly approach. It is also frequently easier to accept criticism from peers than from instructors. Further, peers exchange different perspectives and strategies to solve problems, and construct new knowledge and

meaning through this social interaction (Jarvis, 1987). However, the assessor students also develop new skills as they need to observe and judge their peers' work in relation to standards, which are then transferred to their own performance.

Similarly, in relation to the benefits to the observing students, Martineau *et al.* (2013) compared the learning effects of peers observing each other performing a regular physical examination (n.120), to practicing the same procedure alone (N=65). As a result, those students who observed one another performing the examination learnt more than those who practiced alone ($p < 0.004$). The authors concluded that health sciences educational programs may include opportunities for students to learn from their peers through modelling.

Adding to this study, Nicol *et al.* (2014) evaluated the perception of 82 first-year engineering students on both giving and receiving feedback. A third (27%) stated they learnt from receiving feedback, while the majority (55%) indicated they did learn from both giving and receiving feedback.

1.4.5 Psychometrics of peer-assessment

With regards to peer-assessment scores validity and reliability, the meta-analysis of Falchikov and Goldfinch (2000) found a mean correlation of 0.69 between peer and teacher marks, indicating definite evidence of agreement. Peers' global judgements with well understood criteria resembled more closely teachers assessments than assessing several individual dimensions, and therefore had a higher validity. The authors suggested the peer-assessing of several dimensions or criteria together with an overall judgement, as the optimum approach. Further, an important aspect to increase validity is the students' familiarisation with the peer-assessment criteria. This, together with explaining the students the purpose and goals of the peer-assessment exercise have also been recently stressed (Kamp *et al.*, 2013).

Coming back to the Falchikov and Goldfinch meta-analysis (2000), the authors did not find clear validity differences by subject area, but they highlighted that peers in medically related subjects had a lower tendency to agree with tutors' judgements. Interestingly, this meta-analysis included two studies in dentistry. One on a dental anatomy peer-evaluation (Denehy and Fuller, 1974) and a second one (Jacobs *et al.*, 1975) on self- and peer-evaluation in orthodontics.

The comprehensive review of Topping (1998), found that, in 58% of the studies

(n.18/31) on reliability and validity, peer-assessment was of adequate reliability and validity in a wide variety of applications, while 23% of the studies reported an unacceptably low reliability and validity. Regardless of this, he also found peer assessment to be generally more reliable than self-assessment.

The reliability of undergraduate dental students peer-assessment was evaluated by Satterthwaite and Grey (2008) when 65 pre-clinical dental students' ivory teeth preparations were scored by peers and experienced staff assessors. The data showed no significant differences between scores from both groups ($p = 0.531$). Similarly, another study of the same group also on third-year dental students (Taylor *et al.*, 2013), compared staff grades of 78 typodont full gold crown preparations with peer-assessment marks and measurements from a digital scanning device. The results indicated poor levels of agreement between both staff and peer-assessment marks with the digital mechanism. However, once more, similar levels of agreement were seen between experienced assessors and peer-assessment.

Again in dental education but this time in postgraduate maxillofacial surgeon trainees, Evans *et al.* (2007) studied the criterion validity of peer- and self-assessment scores comparing them with those from trainers. Their results showed that, on average, peer-assessment reflected trainer scores more accurately ($r = 0.83$ for global rating) than self-assessment ($r = 0.55$ for global rating), and that the latter were significantly higher than those given by peers (26.3 versus 28.4, respectively).

Similarly, Sargeant *et al.* (2008) found in a review article, that accurate self-assessment appeared to be difficult and, for some, even impossible. They also suggested that peer assessment may be more accurate than self-assessment and that Multi-Source Feedback (MSF), a workplace-based assessment frequently used in healthcare education and discussed later in this Chapter (page 85), may be used to inform self-assessment. Thus, this combination of peer- assessment to inform self-assessment has for long been recognised to foster reflection on the student's own learning process and learning activities (Dochy *et al.*, 1999).

A recent systematic review on the reliability and validity of 22 different instruments for students' peer-assessment in medical education (Speyer *et al.*, 2011), suggested this practice as an effective format for peer-learning. However, the authors found nothing at all or fragmentary or insufficient data on these instruments' psychometric characteristics. Accordingly, they highlighted that the use of any instrument for

educational purposes can only be justified by its sufficient reliability and validity as well as the discriminative and evaluative purposes of the assessment.

In line with this, and though in a different field of education, Nilson (2003) criticised peer-feedback approaches as having questionable validity, reliability and accuracy and being prone to problems like “blandness, superficiality, inaccuracies, inconsistencies, and so forth”. She ascribed these problems to the fact that peers are being asked to perform a “demanding process of evaluation” instead of just ‘comprehension’ and ‘analysis’ with detailed and specific guidelines. Further, she argued that under these latter conditions, providing and receiving feedback can be very useful in facilitating student learning, and a valuable “life skill”.

Supporting this observation, Ogden *et al.* (2000) used a detailed and simple to use 13-task communication and clinical skills checklist at a summative OSCE station where 125 second-year undergraduate medical students were peer-assessed by Final year dental students while performing a mouth examination. Peer scores were then compared to those from experienced staff dentists. The results showed positive correlations between dental students scores and those from staff members.

A study conducted by Lanning *et al.* (2011), explored the correlation of self-, peer-, near peer- (student of higher course) and staff assessment of second year dental students’ communication skills. Mean scores were for self-assessment 3.86 (sd=0.06); peer-assessment 4.14 (sd=0.04); near peer-assessment 4.07 (sd=0.04); and staff assessment 3.93 (sd=0.10). The highest correlation was observed between peer- and near peer-assessment scores ($r=0.46$, $p<0.0001$). The authors attribute this to the familiarity of the training programme between near peers. Further, the correlation between peer- and staff assessment was negative ($r=0.08$, $p=0.707$). It is arguable that the study did not analyse the possibility of a problem in the design of the rating scale as suggested by Hauser & Bowen (2009), or its compounded items (Mackillop *et al.*, 2011b) that might have appeared confusing specially for novice raters.

Although these and other studies present interesting data regarding the validity and reliability of peer-assessment, it is possible that they overemphasized the comparison between peer and tutor awarding marks and most do not consider the exercise as a learning experience. As posited by Liu and Carless (2006), peer-assessment has the potential to support learning, and ‘measurement’ should not take precedent over ‘learning’.

1.4.6 Students' perception of peer-assessment

Regarding students' perception of peer-assessment, there are studies reporting opinions both in favour as well as against it. Twenty years ago, Williams (1992) evaluated students' attitudes towards new approaches to learning and assessment. Thus, 99 first year business students undertook a self- and peer-assessment exercise and their like and dislike responses were analysed. Ninety percent of the students indicated that they enjoyed and saw benefits in both self- and peer-assessment. The latter included comparison of approaches and of standards and exchange of information. However, these novice students also expressed some drawbacks to the exercise. These were the "criticism of friends" and the "arbitrariness in marking". When asked about improvements, the most frequent answer students provided was "establish firm marking guidelines". It should be noted that this study did not include a students' training in the new assessment approaches.

Similarly, Orsmond *et al.* (1996) also underlined the importance of the marking criteria in biology students' peer-assessment. However, despite tutors' and students' differences in interpreting these criteria, students not only liked carrying out peer-assessment, but also felt the benefits in terms of developing different facets of their learning process. In the same line, third-year geography students expressed their views of a self- and peer-assessment experience as a valuable and enjoyable learning experience which helped them develop skills in independent research, collaboration and communication (Strachan and Wilcox, 1996). In comparison, 96 second-year medical students, who assessed the professional competence of each other, reported that peers identified more strength and weaknesses than they had considered in their own self-assessment, 38% considered these issues to be important enough to become part of their learning plan. Overall, 53% agreed that getting peer-feedback was helpful, while 22% expressed mixed feelings, or disagreement. Sensibly, feedback comments were considered for students to be more helpful than item ratings (Dannefer *et al.*, 2005).

In an exploratory study of students' perceptions on assessment validity, Sambell *et al.* (1997) interviewed different subject students (mainly from social science) and complemented this with other sources of evidence (e.g. nature of tasks and assessment products). They reported the common students' claim that self- and peer-assessment were helping them to develop reflective skills to make judgments about their and other's work. Most students also perceived the benefits of understanding assessment criteria

and to relate it to their own work. However, they also expressed their worries about passing judgments on their friends. Further, some felt threatened by their perceptions of assessing their peers subjectively or a lack of confidence to act as fair assessors. Once again, the authors call for careful and rigorous preparation and support for students to ensure a close match between what students expect of the peer-assessment exercise and the reality of what they actually experience.

Other kind of students' negative perceptions were reported by Beaumont *et al.* (2011) who conducted semi-structured focus-groups to explore the perceptions on peer-assessment of 37 students before entering university. Though some regarded it as constructive and motivational, a much greater proportion expressed bad experiences related to the level of expertise and reliability of peer assessors as well as plagiarism issues.

Sluijsmans *et al.* (2001), conducted a study where 27 university educational science students were asked to summatively rate their peers after a six-week PBL module. They also completed an evaluation questionnaire where 74% of the students were in favour of implementing peer assessment, 73% felt capable of assessing each other, but only 7% felt comfortable when assessing their peers. Though students doubted the reliability of the method, the Generalizability Coefficient reached a high of 0.944 for 13 ratings, which as presented above (Assessment principles in page 45), is considered optimal for even high stakes examinations (Beard *et al.*, 2011). Further, students expressed the need to make room for feedback, the assessment criteria appeared to be difficult to interpret, and students with no prior experience in peer-assessment felt uncomfortable. These last issues underpinned the authors' conclusion, similar to the Williams study (1992), in that there is a need for instruction in peer-assessment in order for students to make reliable judgements.

Interestingly, one year later the same group of researchers (Sluijsmans *et al.*, 2002) reported the effects of peer-assessment training on the performance of student teachers. Fifty students were trained in defining performance criteria, giving feedback and writing assessment reports, whilst 43 made up the control group and received no training. Students with training demonstrated higher quality assessment skill. Further, trained students performed significantly better on subsequent tests than students from the control group. Regarding perceptions, students were significantly more satisfied with the re-designed course. The authors concluded that students could be trained in

assessment skills in order to positively increase their performance.

In another aspect of peer-assessment, 42 undergraduate medical students provided and received anonymous feedback about their presentations on anatomy (Gukas *et al.*, 2008). Results showed that most students were positive and felt comfortable receiving and giving feedback, which they perceived as fair, adequate and helpful, and, interestingly, that receiving feedback made them reflect. However, they also expressed that they would be reluctant to give feedback if anonymity was removed.

In order to avoid the reported peer-assessment and feedback social discomfort and associated responsibility, Burgess *et al.* (2013), investigated students' perceptions of their ability to provide peer-feedback covering both positive and negative aspects of their performance using the positive critique method (Pendleton *et al.*, 2003). This includes four steps as: a) ask what went well, b) tell what went well, c) ask what could be improved, and d) tell what could be improved. Ninety four fourth-year medical students were trained, and they then observed, assessed and provided feedback to each other using this method over a two year period during formative long-case clinical examinations. The majority of respondents (90%) found the exercise a valuable learning experience, including knowledge and skills development, as well as professionalism attributes. However, despite that they found the positive critique method useful as it allowed them to deliver feedback in a standard and professional manner, a significant number (42%) of respondents did not feel confident in providing negative feedback to their peers.

A recent study of before and after peer-assessment process students' perception, on 416 students of 11 different subjects in four fields (Planas Lladó *et al.*, 2013), found that students had a positive view towards peer-assessment both before and after its implementation. They perceived it as a motivating and recommended methodology that facilitated the acquisition of learning at different levels.

This was not the case for 52 first-year electrical engineering undergraduate students who before a peer-assessment intervention were not entirely comfortable or confident in their abilities to assess their peers, despite receiving a thorough training in peer-assessment. After the exercise, however, there was a positive shift overall in both attitudes and confidence (Cheng and Warren, 1997).

Among the very few published studies on dental students' peer-assessment, Larsen and

Jeppe-Jensen (2008) replaced one station of an educational interdisciplinary OSCE by self- and peer-assessment and a later discussion. Sixty-eight third-year dental students who did not receive training on self- or peer-assessment participated in the study. As a result, self- and peer-assessment marks differed widely. Students' opinions and perceptions regarding the benefit of self- and peer-assessment were not uniform, though the majority found it useful.

A recently published study on dental students' perception of peer-assessment after grading their peers' clinical case assignment (Teich *et al.*, 2014), reported that despite participating students (N=55) feeling well prepared for the peer-assessment task, a significant percentage (43.6%) of the group reported that grading the assignment of their peers was not beneficial for their learning process. A possible explanation for these results might be found in the fact that students were not involved in the feedback components of peer-assessment, limiting their role to the measurement (Liu and Carless, 2006).

1.4.7 Limitations of peer-assessment

Despite the overall potential benefits, there remain a number of limitations of peer-assessment. Among the frequently described problems it is not difficult to find reports on "friendship marking", resulting in overmarking, and "collusive marking", resulting in a lack of differentiation within groups (Dochy *et al.*, 1999; Evans *et al.*, 2007; Cushing *et al.*, 2011). On these issues, Papinczak *et al.* (2007) reported a remarkable finding as medical students taking part in a peer-assessment research in Problem-Based Learning, who left prematurely from the study, indicated that friendship marking or lack of honesty was a very important influence on the decision to withdraw. Attitudes among participants included: "I find it difficult to downgrade my peers"; "it is hard to criticize friends"; "no one wants to criticise others in PBL"; "most people are too afraid to honestly mark their peers".

Other less commonly mentioned grading problems are "decibel marking", individuals controlling the highest mark, and "parasite marking", where students fail to contribute but benefit from group marks (Pond and ul-Haq, 1997; Brown and Knight, 1998 p. 59; Dochy *et al.*, 1999). As a way of preventing these problems, the review by Dochy *et al.* (1999) suggested a combination of peer assessment with self-assessment or co-assessment.

Problems in the design and implementation of the peer-assessment exercise can be detrimental for students to learn (Falchikov and Goldfinch, 2000). Peer-assessment conducted in a traditional norm-reference approach, does not have the power to enable students to benefit from their involvement (Falchikov, 2007). This issue has also been cited in the medical education literature (Norcini, 2003a), as norm-referenced scales used to judge peers' workplace performance quality, might prove difficult for junior doctors to use.

Similarly, Arnold (2002) suggested that requesting students to peer-assess each other in all clinical performance dimensions may be subject to a "halo effect" (Gregory, 2004 p. 431) as students might not be able to differentiate between peers' technical knowledge and skills, and professional behaviours. Thus, the author called to limit clinical peer-assessment only to the latter content.

Students do not always accept peer-assessment and feedback as helpful (Beaumont *et al.*, 2011). They have expressed a certain amount of distrust in fellow students' abilities to peer-assess (Planas Lladó *et al.*, 2013), as well as a reluctance to accept any responsibility for assessing or criticising their friends (Dannefer *et al.*, 2005).

The process of implementation of a peer-assessment method needs to be rigorous in order to alleviate and ideally overcome any students' concerns (Sambell *et al.*, 1997). This might explain the reason why some teachers have considered the planning and running of such an exercise to be demanding and time-consuming (Hounsell *et al.*, 2007). However, knowing these issues and incorporating students in the planning process (Strachan and Wilcox, 1996), have been suggested to facilitate and favour peer-assessment implementation (Hounsell *et al.*, 2007).

1.4.8 Feedback

At this point, from the above discussion, it is clear that assessment is an integral component of the learning process. Students' formative assessment results will tell us where they are so that we can support their learning by providing the right feedback on where they should be going (Earl, 2012). Feedback on error then, is a consequence of the assessed performance and is central to students' learning (Black and Wiliam, 1998; Hattie and Timperley, 2007), as it is inevitable that they will have misconceptions that need to be confronted and eradicated; otherwise, mistakes will probably be repeated (Neher *et al.*, 1992; Biggs and Tang, 2011). However, for a long time there has been a

lack of a commonly accepted meaning of feedback (Ramaprasad, 1983; van de Ridder *et al.*, 2008).

Ramaprasad (1983) defined feedback, in behavioural sciences, as information about the gap between actual performance level and the reference level, which is subsequently used to alter that gap. In this 30-year old definition, he already pointed out, as discussed below, the need for a feedback that is meaningful, understood and that is acted upon.

Latter, Hattie and Timperley (2007), defined feedback as “information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one’s performance or understanding. A teacher or parent can provide corrective information, a peer can provide an alternative strategy, a book can provide information to clarify ideas, a parent can provide encouragement, and a learner can look up the answer to evaluate the correctness of a response”. Despite mentioning ‘parents’ as a possible ‘agent’ provider for feedback, which could be interpreted as a definition targeted for school students, it can also be understood that feedback can have different purposes: corrective, alternative, clarifying, and encouraging. As discussed below, this is extremely important as these and other signs, or moderators, will define the usefulness of the feedback.

A recent feedback definition in higher education offered by Boud and Molloy (2013), states that it “is a process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standard for any given work, and the qualities of the work itself, in order to generate improved work”. Widening its meaning, Evans (2013) includes exchanges occurring beyond the immediate learning context, actively and/or passively sought and/or received and from a range of sources. This broader approach is particularly important for the evaluation of healthcare students and professionals through multisource feedback (Wright *et al.*, 2012), as described in the next section. In medical education, feedback has been defined as “a way in which learners become aware of the gap between their current level of knowledge or skill and the desired goal” (Wood, 2011).

In spite of explaining the concept, not all these definitions mention the need for the feedback receiver to take an action for learning to happen (Sadler, 2010), or the level of students’ engagement (Evans, 2013), or about the quality of the feedback information, particularly its promptness, tone, amount and level of detail, style, clarity, structure and relevance (Wingate, 2010; Nicol *et al.*, 2014). Consequently, once again, feedback today is not just telling the students where and what to improve (Bloxham, 2009); they

must play an active role during (Carless, 2006) and after the process (Nicol, 2010). Further, as presented by Boud and Molloy (2013), the current duty of feedback is not just to improve performance ‘now’, but also on the capacity of the learner to better manage future and different tasks.

This reconceptualization of feedback in higher education, has come as a response to the learners’ complaint that they never receive feedback (Branch and Paranjape, 2002) or that it is not enough (Boud and Molloy, 2013), which has consistently been reflected in the low levels of students’ satisfaction with feedback in the National Students Survey (NSS) in the UK (Higher Education Funding Council for England, 2011 pp. 8-14). The NSS is undertaken by final year undergraduate higher education students in the UK, and is recognised as an important measure of student satisfaction (National Student Survey, 2014). Results from the NSS are publicly available (Unistats, 2014), allowing comparison of different universities and courses. Thus, it is within every university’s best interests to achieve the highest NSS scores as possible (Holmes, 2014).

Accordingly, an analysis of the report of the findings and trends of the sixth annual NSS carried out in 2010, published by the Higher Education Funding Council for England (2011 p.42), as well as the available data from Unistats (Unistats, 2014), allowed comparison of the level of students’ satisfaction in those three questions related to feedback for medical and dental students (as both subjects are presented together), those following education studies, and the global score (Table 1-2).

Table 1-2 Percentage of UK NSS higher education full time student’s satisfaction (“mostly agree” and “definitely agree”) to the three questions related to feedback (numbers 7, 8 and 9) at three different years (2007, 2010, 2013), for students from medicine and dentistry, education, as well as the overall score. Source: Higher Education Funding Council for England (2011 p.42) and Unistats (2014).

NSS Respondent Satisfaction				
Question		Medicine & Dentistry	Education	Global score
7. Feedback on my work has been prompt	2007	39 %	56 %	53 %
	2010	40 %	64 %	58 %
	2013	59 %	73 %	67 %
8. I have received detailed comments on my work	2007	31 %	72 %	59 %
	2010	33 %	74 %	62 %
	2013	51 %	81 %	72 %
9. Feedback on my work has helped me clarify things I did not understand	2007	38 %	60 %	53 %
	2010	41 %	64 %	57 %
	2013	54 %	73 %	67 %

From the information shown in this table it is possible to highlight the following:

- i. Medical and dental students show consistently lower satisfaction scores compared to education students and the global score.
- ii. Education students display higher satisfaction scores than the global score.
- iii. There is a gradual improvement in all three questions for medicine and dentistry, education and the global score, across the 2007 to 2013 period.
- iv. Medicine and dentistry made the bigger improvements between 2010 and 2013 while education has done it more gradually from 2007.
- v. Despite the improvements, there are still more than 40% of future medical doctors and dentists who have the perception that they are not receiving prompt feedback.
- vi. From these medicine and dentistry students, 46% do not see the received feedback as helpful.

As described in the literature, the natural response to the initial low level of feedback satisfaction among higher education students, has been an extra effort to enhance the quality of the feedback information provided by teachers, which appears enormously resource-heavy (Bloxham and Campbell, 2010; Nicol *et al.*, 2014).

Despite the fact that feedback satisfaction in the NSS has increased, there is much work to be done. This because in spite of these better NSS figures, as a result of the described attempts to develop teacher feedback, recent studies show that students' learning, as a result of the provided feedback, does not improve (Orsmond and Merry, 2011; Evans, 2013; Orsmond *et al.*, 2013).

In this regard, Boud and Molloy (2013 pp. 7-8) have highlighted three conventional assumptions that need to be challenged in order to understand and improve the current feedback provision to enhance students' learning. These are:

- i. Feedback constitutes one-way flow of information from a knowledgeable person to a less knowledgeable person.
- ii. The job of feedback is complete with the imparting of performance related information.
- iii. A generic model of best-practice feedback can be applied to all learners and all learning situations.

Though these statements are very much interrelated, they will be approached separately for didactic reasons.

First, feedback must ‘not’ constitute a one-way flow of information from a knowledgeable person to a less knowledgeable person. Indeed, scholars now believe that, if feedback is to enhance students learning and develop expertise, it needs to be reconceptualised as a dialogue (Nicol, 2010), a process where students must play an active role (Sadler, 2010), whereby they are given opportunities to construct their own meaning from the received feedback. Thus, students need to engage (Price *et al.*, 2011) with the feedback message, take it, analyse it, ask questions about it, discuss it with others as an ongoing socially-embedded process (Price *et al.*, 2011), and finally, connect it with prior knowledge (Carless *et al.*, 2010; Nicol *et al.*, 2014). The concept of ‘feed forward’ has been introduced to explain this process (Sadler, 1983), on how to improve the students' use of tutors' comments by moving feedback forward. Further, the three phases model to “move feedback forward” later developed by Sadler (1989) can be combined with the Hattie and Timperley (2007) “model of effective feedback” in the following steps: a) According to Sadler (1989), the student needs to start by knowing the standard, goal and criteria for their desired degree of performance or excellence, which in the words of Hattie and Timperley (2007) corresponds to the answers to tutor or peer questions as where am I going? and, what are the goals?, known as ‘feed up’; b) Then, the student must recognise how her or his current performance relates to this standard or criteria, by asking her- or him-self how am I doing?, what progress have I made toward the goal?, recognised as ‘feed back’; c) Finally, the student needs to understand and design a strategy to close the gap between her or his current performance and the goal, by asking where to next?, what activities do I need to undertake to make a better progress?, which corresponds to the notion of ‘feed forward’.

A similar process called the “feedforward interview” has been described by Kluger and Van Dijk (2010) for use in the clinical setting. Surprisingly, none of these studies reference each other.

In essence, by providing ‘feed forward’ feedback, students’ immediate needs of the piece of work or performance are satisfied, delivering high quality information, but at the same time, by encouraging them to engage with the message, they start monitoring their own work to become self-regulated learners (Nicol and Macfarlane-Dick, 2006). This process has been also coined as “sustainable feedback” as it supports students in

self-monitoring their own work independently of the tutor (Carless *et al.*, 2010). Further, as discussed above (Assessment purposes in page 43), fostering students' metacognition and self-regulation is also central to the concept of assessment *as* learning (Earl, 2012), where students become active critical thinkers, comfortable with reflection, making sense of the feedback message, relating it to prior knowledge and using it to construct new learning.

However, in the current educational environment where resources to hire new staff are limited and the number of students are increasing (Light *et al.*, 2009), there arises the question of how all these demanding jobs are going to be accomplished (Sadler, 1989; Price *et al.*, 2010). Thus, some scholars have suggested the incorporation of peers to support the feedback process (Boud *et al.*, 1999). This is informed by research showing that peer-feedback's accessible language is often better understood than that of the tutors' (Topping, 1998; Boud and Falchikov, 2007). Further, receiving feedback from multiple peers has been reported to be highly reliable (Cho *et al.*, 2006), and more helpful in improving the quality of students' assignments than receiving feedback from one peer or a single expert (Cho and MacArthur, 2010). Moreover, recent research has shown that formative peer-assessment may be a vehicle to close the gap between the feedback *given* to students and feedback effectively *used* by them (Cartney, 2010), and that repeated involvement in dialogic peer-feedback interactions with a self-assessment component would support the process of self-regulation (Carless *et al.*, 2010). On this point, peer-feedback provides a more social learning format, taking learning out of the private domain (Liu and Carless, 2006), which has been seen to achieve higher standards in comparison to the one-way flow practice of tutors telling students about the quality of their work (Sadler, 2010).

Second, the job of feedback is 'not' complete with the imparting of performance related information. As mentioned above, students need to take action on the received feedback to learn from it. It can be accepted that the process of feedback might be prompted by a tutor's words or writing, but the process is not concluded until action by the student occurs (Boud and Molloy, 2013).

Thus, aiming to explore the extent to which undergraduate social work students acted on their assessors' feedback, Crisp (2007) concluded that providing feedback alone was not sufficient to improve students' written work. She also noted that assuming that students understood the received feedback because they had not sought subsequent

clarification about comments made, was wrong. In line with this, Higgins *et al.* (2002) focused on business and humanities students' understanding of feedback, highlighting that students need to have misconceptions 'explained' as soon as possible, and that simply telling them their mistakes, was not enough. Further, a recent Cochrane Collaboration review on the effects of audit and feedback on professional practice and healthcare outcomes, found that when feedback was given both verbally and in writing, and when it includes a clear target and an action plan, it is most effective (Ivers *et al.*, 2012). This reaffirms the conclusion that feedback should not be just 'giving' information, and that it crosses the discipline as well as the undergraduate and professional borders.

Casting doubts on whether the practice of feedback as just 'telling' is ever to change, Bailey and Garner (2010) interviewed 48 lecturers from across departments and faculties, who stated that despite their university's commitment to ensure timely and useful formative feedback on assignments, this was not having the intended effect. Lecturers were uncertain about what use students made of the provided feedback, and honestly but worryingly, many seemed to have become "indifferent to the educational value of written feedback" and did it just to comply with institutional policies.

This apathy might be explained by the findings of Evans (2013) in that the demands on the lecturer to support students' access to and engagement in feedback are 'huge'. Notably, the lecturers in the former study made clear their doubts of the use students gave to the provided feedback, limiting their role to the provision of feedback, which might explain why students still feel, as expressed in the NSS, that feedback on their work has not helped them clarify things they did not understand. In line with this, Orsmond and Merry (2011) suggested a misalignment in feedback provision as developmental aspects of students' learning were rarely addressed in tutor feedback. Further in this dissonance, Carles (2006) established that tutors believe that they are providing more detailed feedback and that it is more useful than students do, while Adcroft (2011) found that students perceive they receive feedback much less frequently than tutors perceive they give it.

Ideally then, and following the concept of assessment *as* learning (Earl, 2012) described above, lecturers should do their best to 'engage' students in feedback (Evans, 2013), going beyond the judgement of 'where' and 'what' to improve (Bloxham, 2009), by helping them to understand the meaning of the feedback by having a dialogue (Nicol,

2010; Orsmond *et al.*, 2013) and encouraging them to take action on it (Sadler, 2010). This should also consider other factors such as tutor's low credibility among students as a reason for their unwillingness to act on their tutor's feedback (Orsmond *et al.*, 2005).

However, not all the blame is on the tutors. In the study by Wingate (2010), it became apparent that students paid little attention to their tutors' feedback comments and did not act upon them. The reasons for this included their low motivation based on the enjoyment of the degree programme they were following, and the self-perception of their ability as writers. The author also highlighted the need, as feedback providers, to pay more attention on the comments given, targeting them differently to high- and low-achieving students. This implies knowing the students better, leading us to the third Boud and Molloy (2013) challenged assumption that one feedback model suits all learners and situations.

Third, a generic model of best-practice feedback can 'not' be applied to all learners and all learning situations. The reasons for this are that university-based and workplace environments have variations in contexts, persons and risks, which do not allow for an appropriate application of currently advocated feedback models. Further, the increasing numbers and diversity of university students coming from a wide range of educational experiences are being educated for increasingly diverse practices and workplaces. This does not allow feedback to be based on a common set of assumptions, as was done before (Boud and Molloy, 2013).

A reported quick-fix solution to the above presented NSS low scores for feedback, in which tutors were encouraged to signal and underline their use of anything that seemed like feedback on every occasion they could think of, does not recognise the magnitude of the real problem and the need for change (Boud and Molloy, 2013). These approaches will probably not satisfy students because what they are seeking is a dialogic individual communication (Nicol, 2010; Dowden *et al.*, 2011).

It should also be noticed that quick feedback tricks which are not contextualised may have no effects or even negative consequences, as they should consider the task, the process, the learning setting, the learner's motivation and self-regulation (Kluger and DeNisi, 1996; Hattie and Timperley, 2007; Kluger and Van Dijk, 2010). For example, Kluger and DeNisi (1996), found in their meta-analysis that feedback effectiveness decreases as attention moves up the hierarchy closer to the self and away from the task motivating the feedback.

It is important to ensure the feedback message is targeted at students at the appropriate level (Hattie and Timperley, 2007). Thus, feedback is not uniform in use and in concept, and when, how, and by whom feedback is delivered matters and the effects are variable as a function of the specificities of the situation (Eva *et al.*, 2012).

1.4.9 Feedback that works

There are many claims that learning and performance can be optimised by the provision of feedback. In the Hattie and Timperley (2007) original review, the effect size of feedback from 12 meta-analyses including 196 studies was $d=0.79$ (twice the average effect of $d=0.40$). In a recently published book, Hattie (2012) upgraded the factors achievements by including now over 900 meta-analyses, where feedback is placed in the top ten influences on achievement, though with considerable variability.

Consistently, another meta-analysis of 131 studies conducted by Kluger and DeNisi (1996), addressed the effect of 36 feedback moderators and found an average effect size of $d=0.38$. The wide range included a maximum effect size of $d=0.55$ for “task feedback about changes from previous trials” as well as in feedback for “not complex tasks”, to a negative effect size of $d=-0.31$ for task feedback designed to “discourage the student”. Similarly, there were ineffective feedback interventions such as ‘praise’ for task performance, and with a weak effect such as physical tasks (negative effect) and following rules tasks. The authors highlighted that most of the variances could not be accounted for by sampling error. Surprisingly, 32% of the feedback interventions had a negative effect size, though only ‘discouraging feedback’ was statistically significant.

In addition, the Cochrane Collaboration review on the effects of audit and feedback on professional practice and healthcare outcomes, referred above, found three studies with large negative effects related to prescription of medicines and laboratory test utilization (Ivers *et al.*, 2012).

Smaller studies on students’ experiences have also helped in identifying feedback perceived as unhelpful to improve learning. Weaver (2006) found that comments which were too general or vague, lacked guidance, focused on the negative, or were unrelated to assessment criteria, were not beneficial.

A systematic review on assessment, feedback and physicians’ clinical performance conducted by Veloski *et al.* (2006), found that 32/41 (74%) studies that evaluated the independent effect of feedback on physician performance (baseline measurement of

performance, feedback and follow-up assessment), demonstrated a positive impact. The most effective feedback on physician performance was found to be provided by a credible and authoritative source over a number of years. Unfortunately, as the authors stated, the variation in outcome variables precluded any systematic analysis of effect sizes.

From these and other studies, it appears that feedback is more effective when:

- It directs information to enhance self-efficacy in the task and to more effective self-regulation, promoting reflection on actions (Archer, 2010), and stimulating students' motivation (Wingate, 2010). In this case, feedback is likely to produce remarkable performance improvements (Kluger and DeNisi, 1996). Therefore, students need quality over quantity feedback (Dolmans, 2013) about their performance against clearly defined learning outcomes (Harden and Laidlaw, 2013). This is particularly important for the disadvantaged and low-attaining learners, as they benefit the most from a formative assessment feedback (Black and Wiliam, 1998).
- It is phrased in as non-evaluative language as possible. Comments like their performance was 'totally inadequate' are not helpful (Harden and Laidlaw, 2013). It should always be remembered that trainees are generally apprehensive about feedback and their fear will only disappear when they realise that feedback is essential for effective learning and the development of competence (Pelgrim and Kramer, 2013).
- It is given in a timely manner (Subramanian *et al.*, 2013), when memory is still fresh for both observer and trainee (Weaver, 2006; Kilminster *et al.*, 2007; Shute, 2008), but allows emotions to be reduced (van der Leeuw and Slootweg, 2013). Harden and Laidlaw (2013) found that providing feedback immediately following an Objective Structured Clinical Examination (OSCE) was a useful and powerful learning experience. Further, prompt feedback has been suggested as a condition for a clinical behaviour to be corrected (Ramani and Krackov, 2012).
- It is provided at the end of the task (terminal feedback) (Walsh *et al.*, 2009), in a private setting (Neher *et al.*, 1992), giving the students some time to reflect on their learning (McMillan, 2011).
- It builds on changes from previous trials containing cues that support learning, drawing attention to discrepancies between the task and the standard (Kluger

and DeNisi, 1996). Thus, Wingate (2010) found that students who had utilised their earlier feedback comments improved in those areas previously criticised.

- It includes an action plan with personal goals and targets that are specific, clear and challenging but overall task complexity is low. Kamp *et al.* (2013) observed that the impact of peer-feedback at the ‘process’ level according to Hattie and Timperley model (2007) (explained below in this section), can be increased by combining it with individual reflection by means of goal setting with face-to-face discussion.
- It moves away from a trainer-driven monologue to a valuable dialogue (Archer, 2010; Orsmond *et al.*, 2013), where feedback, in a socio-constructivist paradigm (Vygotsky, 1978; Jarvis, 1987), is seen as facilitative, enabling students to make their own revisions and, through dialogue, help students to gain new understandings without dictating what those understandings will be (Evans, 2013). This is hindered by the classic medical hierarchical model (Archer, 2010), in which the cognitivist perspective is associated with a directive telling approach where feedback from an expert provides information to a passive novice (Evans, 2013).
- The feedback provider is a supervisor, peer (Ivers *et al.*, 2012), or a role model near-peer (Nelson *et al.*, 2013) who is aware of the qualities of the discussions that result in successful feedback (Mehta *et al.*, 2013), and that can separate the content from the relationship (van der Leeuw and Slootweg, 2013). When provider and trainee arrange and plan the observation and feedback, a greater benefit can be expected (Pelgrim *et al.*, 2012a). Further, students’ receptivity of the feedback is influenced by the credibility of the provider (Watling *et al.*, 2012a). Accordingly, they will reject feedback from a not credible assessor, or not truly engage in the creation and exchange of informed and accurate feedback (Watling *et al.*, 2008).
- It is provided frequently (Kilminster *et al.*, 2007; Ivers *et al.*, 2012), in stress-free conditions (Rolfe and McPherson, 1995), ensuring all performance evidence is available and first-hand (Harden and Laidlaw, 2013). However, feedback provision in busy, time-constrained clinical settings is frequently overlooked (Archer, 2010; Sabey and Harris, 2011).
- It is given both verbally and in writing (Ivers *et al.*, 2012) in a comprehensible language to the student (Higgins *et al.*, 2002), and in a way that facilitates the development of self-assessment (reflection) in learning (Nicol and Macfarlane-

Dick, 2006).

- There are perceived low rather than high levels of threat to self-esteem, because low-threat conditions allow attention to be paid to the feedback. It should also encourage positive motivational beliefs (Nicol and Macfarlane-Dick, 2006).
- Combined with other interventions such as effective instruction in classrooms (Kluger and DeNisi, 1996), reminders (Ivers *et al.*, 2012), educational programmes and practice guidelines (Veloski *et al.*, 2006).

In order to analyse these factors in more detail, the Hattie model of feedback will be used as a framework to identify conditions to maximise the positive effects of feedback on learning (Hattie and Timperley, 2007; Hattie, 2012). Before going deeper in the model, it should be stated that a broader aim is to create, as called by Sargeant *et al.* (2011) a “culture of improvement” where sharing and seeking feedback to improve performance is the norm of work and learning.

Thus, the rationale of Hattie’s feedback model is to reduce discrepancies between current understanding and performance and a given goal. However, from the information presented above, we already know that not all strategies to reduce this gap and enhance learning are equally effective. Accordingly, the first notions to start with in Hattie’s model of feedback are the answers to three major questions asked by a tutor or peer, as previously presented. By recapping, these are:

- a) Where am I going? (‘feed up’ stage)
- b) How am I going? (‘feed back’ stage)
- c) Where to next? (‘feed forward’ stage)

The answers to these questions can be very effective to reduce the gap, but can also be otherwise. This depends on the ‘level’ at which each of the three questions is focused. According to Hattie and Timperley (2007), these levels are:

- 1) ‘Task and product’
- 2) ‘Process’
- 3) ‘Self-regulation’
- 4) ‘Self’

These four levels are discussed below and where appropriate, they are complemented with evidence from other studies.

Feedback about the ‘task and product’ level is also known as ‘corrective feedback’ or ‘knowledge of results’. It can be very powerful in novice learners, and when it is more information focused (e.g. correct or incorrect responses), leads to the students to get and provide more or different information relevant to the task, and building more task knowledge. It corresponds to the answer(s) how well has the task been performed? Is it correct or incorrect? It is not only critical, but is the foundation to build effective ‘process’ and ‘self-regulation’ levels (Hattie, 2012). However, the learner needs an explanation as to what she or he did or did not do to meet the expectations. Simply marking an examination (Rolfe and McPherson, 1995; Shute, 2008) or telling the students what went right or wrong, without an explanation is less likely to improve their performance (Harden and Laidlaw, 2013).

For this to occur in the clinical setting, the ‘task’ performance must be observed. This makes a big difference, as if the feedback is to be given about a written ‘task’, the tutor or peer feedback provider does not need to observe the procedure, but its result. Unfortunately, performance observation in health education is low (Williams *et al.*, 2003; Kogan and Hauer, 2006; Norcini, 2011), and, as discussed below (Clinical workplace-based assessment and the role of peers in page 81), high proportions of students frequently report difficulties in finding a clinical assessor (Quantrill and Tun, 2012).

The second level focuses the tutor or peer feedback to the ‘process’ used to complete the task. It looks for the student to understand and connect ideas, providing strategies to identify mistakes and transfer the learning experience from these mistakes to other more difficult tasks. This move from ‘error detection’ to ‘error correction’ is highly desirable and has been related to student motivations to close the gap between their current performance and the goal (Boud and Falchikov, 2007). The ‘process’ feedback that contains specific narratives is then more effective to enhance deep learning and gets higher satisfaction scores from recipients (Overeem *et al.*, 2010), as it is more readily integrated and assimilated (Sargeant *et al.*, 2008). The ‘error correction’ with frequent, specific and precise narratives, is particularly important for those novice students who cannot judge their performance, or competency, accurately (Hauser and Bowen, 2009), as they cannot learn by simply observing their tutors (Hendricson and Kleffner, 1998).

Whilst Grieveson *et al.* (2011b) reported the high perception of dental trainees in that feedback received was provided in a supportive way, had a positive effect on their

training, encouraging them to be reflective in their clinical practise, not all feedback practices reach this level. A study by Mutch (2003) in a university business environment, concluded that feedback with no indication of how deficiencies could be addressed was of little value as it could only be used by those students who recognise what the implied developments were. In line with this, a big study including 460 staff and 1740 students in the eight publicly funded universities in Hong Kong (Carless, 2006), found that 38.4% of the tutors thought students were often given detailed feedback which helped them improve their next assignment, whilst only 10.6% of students responded in the same way. Further, 37.8% of students felt that feedback was rarely followed by actions to improve student learning, as opposed to 16.1% of tutors.

The third feedback level focuses on the student ‘self-regulation’, or, in other words, to the student’s monitoring their own learning process towards the goal. The aim here is to develop the student’s capability of internal feedback or self-assessment, as when they do, they reduce discrepancies between where they are in their learning and the desired outcomes, more effectively. For example, feedback at this level can be focused to help the student to self-evaluate by asking reflective or probing questions on ‘when’, ‘where’ or ‘why’ related to the task. Students who learn the metacognitive skills of self-assessment, develop their self-appraisal (facility to review and evaluate their competences through reflection on action) (Archer, 2010) and self-management (monitoring and regulating their behaviour through planning, error correction, and fixing strategies) skills as well, and by doing so, they know how and when to seek and receive feedback from others. As discussed above (Reported benefits of peer-assessment in page 54), many educational initiatives include in their programme the combination of a self- and a peer-assessment component (Falchikov, 2007) aimed at helping students to take more responsibility for their own learning (Black and Wiliam, 1998). Further, McMillan (2011) suggests asking the student for a self-appraisal of their performance – identify aspects that went well, areas of difficulty and possibilities for change - before offering feedback in order to develop their self-reflective skills.

The fourth level is feedback directed to the ‘self’ as a person. It is important to be aware that personal feedback, commonly considered as ‘praise’, such as ‘you are a great student’ or ‘great effort’, often directs the attention away from the task, process, or self-regulation. Precisely because of this, praise and feedback about the learning should be kept separate. Praise includes little information about the task performance and therefore does not help in answering the three feedback questions. Thus, whilst Skipper

and Douglas (2012) found that ‘self’ feedback was likely to have a negative effect particularly in those students who are not successful, Boehler *et al.* (2006) reported that students who received general compliments (praise) provided higher satisfaction rating compared to those who received feedback. However, only the latter group increased their performance. Further, Parkes *et al.* (2013) tested the effectiveness of the frequently used (Kogan *et al.*, 2012) “Sandwich Feedback Technique” – Praise; Critique; Praise - (Davies and Jacobs, 1985) in a peer-feedback exercise of 3rd year medical students, and found that students believed feedback sandwiches positively impacted their subsequent performance, but in reality they did not. Feedback should not undermine self-esteem, but should not simply consist of praise (Kluger and DeNisi, 1996).

An interesting point raised by Bing-You and Trowbridge (2009) is that the current students’ dissatisfaction with feedback (referred to medical students in the USA), “may reflect a greater desire for praise than for constructive information to help them learn”. However, moving away from ‘perceptions’ and focusing on students’ learning, feedback needs to move from the task (what do I know and what can I do?) towards the process or understandings necessary to learn the task (what do I not know and what can I not do?), and from here to self-regulation about continuing beyond the task to more challenging goals (what can I teach others and myself about what I know and can do?) (Hattie, 2012). However, in the medical education literature there still are calls to combine both compliments and feedback (Boehler *et al.*, 2006).

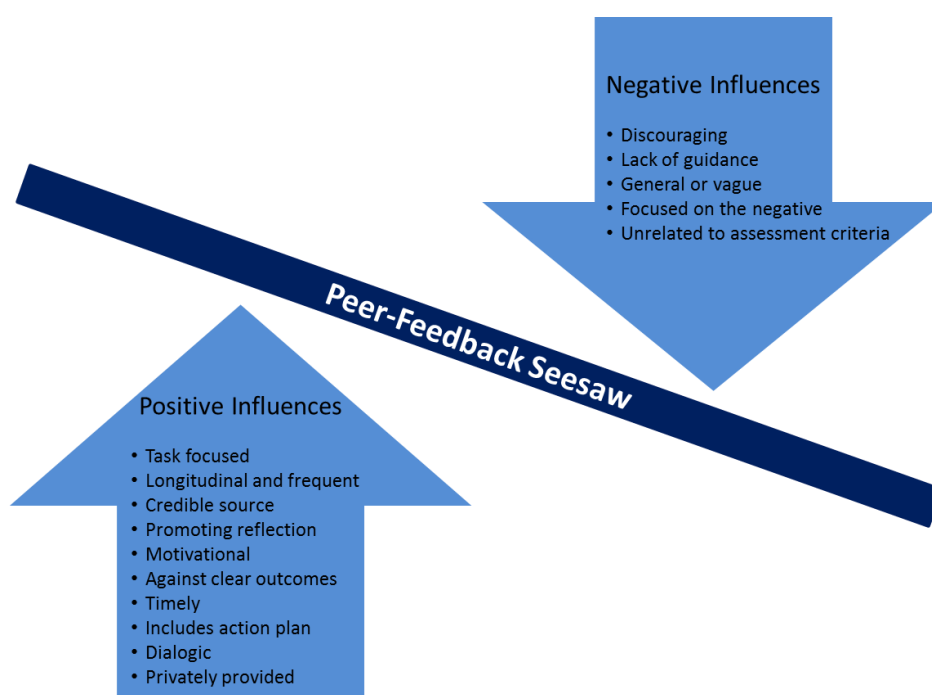


Figure 1-6 Graphical interpretation of the peer-feedback characteristics that could have a positive or negative influence. This could be used for peer-assessment training purposes.

Some authors have stated that a regular self-, peer- and/or multisource assessment-feedback strategy (Aronson, 2011), with students actively working in different roles as part of a team (McMillan, 2011), facilitates and is crucial to deeper learning (Boursicot *et al.*, 2011; Manogue *et al.*, 2011; O'Donnell *et al.*, 2011). In line with this, Sweet *et al.* (2008) perceived that students who worked together and engaged in discussions about what they do in the clinic, can be expected to develop a deeper learning with changes in their perspectives.

1.4.10 Reflection and learning

Reflection, reflective learning, reflective writing and reflective practice are being increasingly used not only in higher education but also in professional development (Mann *et al.*, 2009; Gonzalez *et al.*, 2013). However, there are some differences in the views of how reflection is defined (Sumsion and Fleet, 1996), though consistent with the Latin origin of the word as “to bend” or “to turn back” (Moon, 2007). A few selected definitions of reflection are:

John Dewey (1938), philosopher, psychologist, and educational reformer.

“An active, persistent and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and further conclusions to which it leads...it includes a conscious and voluntary effort to establish belief upon a firm basis of evidence and rationality’ and the further conclusion to which it tends”.

David Boud *et al.* (1985 p. 19), researcher and teacher of adult higher and professional education.

“A generic term for those intellectual and affective activities in which individuals engage to explore their experiences in order to lead to a new understanding and appreciation. It may take place in isolation or in association with others”.

John Sandars (2009), medical doctor, lecturer in medical education.

“Reflection is a metacognitive process that occurs before, during and after situations with the purpose of developing greater understanding of both the self and the situation so that future encounters with the situation are informed from previous encounters”.

Jennifer Moon (Moon, 2013), educational developer and professional trainer.

“A form of mental processing - like a form of thinking - that we use to fulfil a purpose or to achieve some anticipated outcome. It is applied to relatively complicated or unstructured ideas for which there is not an obvious solution and is largely based on the further processing of knowledge and understanding and possibly emotions that we already possess”.

From these definitions, it can be noticed that Boud *et al.* (1985) and Sandars (2009) are focussed on the process of reflection, while the latter adds the temporal component stating it can occur “*before, during and after*”, as well as the prospective use of the outcome. Similarly, Moon (2013) mentions the purpose of reflection to anticipate an outcome, and as Boud *et al.* (1985) do, she considers emotions. Finally, Dewey (1938) presents a holistic view of the process of reflection for learning (Eraut, 1995), highlighting its voluntary character.

On a further analysis of these definitions, it can be notice that they do not make an explicit difference between the previously described Schön (1983) concepts of reflecting during the event (reflection-in-action) or after the event (reflection-on-action) (described in Reflective practice learning in page 35). Probably the definition by Sandars (2009) is the only one clearly focusing on reflection-on-action, when he mentions ‘turn back’ thoughts. Though debatable, this is in agreement with other medical researchers who have stated that in “medical education”, most reflection is on-action (Aronson, 2011), if at all (Branch and Paranjape, 2002).

This might be what in reality happens during medical education. That is, exclusively adopting the previously discussed Kolb’s model (1984) where students only reflect ‘after’ their experiences. As Finger and Asun (2001 p. 45) put it, they are reflecting following a ‘mistake’, that is, “learn by trial and error”. However, this approach would hinder students’ clinical reasoning since this is conceptualized as reflection-in-action (Schön, 1987). Further, as reflection-in-action involves reflecting and doing it in the midst of action (Stegeman *et al.*, 2013), “it implies that the professional has reached a stage of competence where she or he is able to think consciously about what is taking place and modify actions virtually instantaneously” (Hatton and Smith, 1995). Supporting this approach, Mamede *et al.* (2007), found that complex and unusual clinical cases led medical residents to switch from automatic (non-analytical reasoning)

to reflective reasoning.

It should be noted that this is not a call to abandon Kolb's learning cycle; on the contrary, this is a request for a cultural change to include in the teaching of our students' a requirement for them to reflect 'while' practicing and 'after' practicing. As presented by Boyd (2008), dental practice needs both reflection-in-action and reflection-on-action, to prepare students for a lifetime of professional development (Frankford *et al.*, 2000). This entails the creation of new ways of thinking and acting about practice problems, where the feedback process has been mentioned as key to successful growth and learning (Lee and Caffarella, 1994). In the same line, other researchers have called for encouragement of reflection-in-action in health professional education in order to achieve truly effective feedback (Archer, 2010).

The need for reflection to accomplish effective learning was emphasised by Dewey some time ago (Dewey, 1909; Dewey, 1938), not forgetting the ancient discussions of Plato (Plato, 1952 pp. 684-685 [720]) regarding the "thoughtful diagnosis and reasoning about treatment and explanation to the patient" (Ericsson, 2008) by a freeman doctor. As well as the previously described models of experiential learning (Kolb, 1984) and reflective practice (Schön, 1983) (pages 33 and 35, respectively), Mezirow's (1991) constructivist model of transformative learning, also encourages a reflective interpretation of experiences to "make meaning" of them, as fundamental to what learning is all about.

Despite the known benefits (Branch and Paranjape, 2002; Mann *et al.*, 2009) and the requirements to incorporate reflective teaching (American Commission on Dental Accreditation CODA, 2010; Australian Dental Council, 2010; General Dental Council, 2012a), its implementation is challenging (Boud and Walker, 1998; Grant *et al.*, 2006; Mann *et al.*, 2009; Sargeant *et al.*, 2009), in dental traditional curricula (Pee *et al.*, 2002; Woodman *et al.*, 2002; Ashley *et al.*, 2006). Further, the reality seems to be that while feedback to promote reflection is not used often enough, reflection is probably used even less (Branch and Paranjape, 2002), despite students' recognition of the need for reflection especially after starting clinical practice (Carr and Johnson, 2013).

In the field of healthcare education and practice, reflection on one's own experiences is generally recognised as a core, critical competence (Mann *et al.*, 2009; Lindström *et al.*, 2011) and the foundation of purposeful learning (Amulya, 2004). However, as Jarvis has pointed out (Jarvis, 1987 p. 84), some individuals do not learn from experiences.

Precisely on this, a key finding in Tsang and Walsh (2010) study on last year oral health students, was that prior to implementing a clinical reflective learning programme, “clinics were attended and then forgotten”. It is today accepted that knowledge and practical skills alone are not enough to become a medical professional (Boenink *et al.*, 2004), as the critical analysis of lived experiences, in order to understand their broader context, allows the learner to actively integrate the new resulted knowledge (Moon, 2013). By developing the habit of exploring and being curious about our own struggles, uncertainties, and breakthroughs, we open the possibility of a purposeful learning, not from external sources, but from our own work and lives (Amulya, 2004).

A number of studies explain the effects of reflection on the learning processes of students and trainees. For example, reflection has been reported to improve, develop or facilitate a range of attributes, including critical thinking (Phan, 2008; Mitchell *et al.*, 2009) clinical reasoning skills (Baernstein and Fryer-Edwards, 2003; Sandars, 2009), diagnostic reasoning abilities (Sobral, 2000), particularly in complex and unusual cases (Mann *et al.*, 2009), medical-humanism skills (Epstein, 1999; Wiecha *et al.*, 2002; Gracey *et al.*, 2005), as well as technical skills, evidence-based decisions (Epstein, 1999) and professionalism (Mofidi *et al.*, 2003; Stern and Papadakis, 2006; Phan, 2008).

Pedagogically and from a constructivist viewpoint (Knowles *et al.*, 2012), reflection allows the integration of new learning into existing knowledge and skills (Mann *et al.*, 2009), promoting self-regulated learning activities (Boud *et al.*, 1985; Grant *et al.*, 2006) which assist and encourage professional development (Tsang and Walsh, 2010; Lindström *et al.*, 2011). This is accomplished by increasing the self-awareness of students and trainees in the learning processes (van den Boom *et al.*, 2007; Sandars, 2009), and so enhances the learning opportunity (Sobral, 2000; Lonka *et al.*, 2001; Cleary and Sandars, 2011) through a deeper approach to learning (Sobral, 2005; Sandars, 2009; Tsang and Walsh, 2010).

By way of contrast, there are some studies that report no measurable academic improvement after reflective writings (Lew and Schmidt, 2011), or problems with, or adverse effects from, forced reflection. Thus, an “instrumental or rule-following approach to reflective activities” has been suggested as leading to reflection without learning and inappropriate levels of disclosure with strategic responses, tension between public and private reflections, hostility and even moral questions (Boud and Walker,

1998; Boud, 1999; Hobbs, 2007; Vivekananda-Schmidt *et al.*, 2011), as well as the necessity (Jonas-Dwyer *et al.*, 2013), with its associated anxiety to find time for reflection (Burnard, 1995; Dornan *et al.*, 2002; Pearson and Heywood, 2004).

Reflective learning does not happen intuitively, spontaneously, voluntarily, or even devoutly (Grant *et al.*, 2006; van den Boom *et al.*, 2007; McDonald, 2010; Chambers *et al.*, 2011), especially in the case of new generations of students (Sandars and Homer, 2008; Wald *et al.*, 2012). Some students believe that learning the technicalities of practice are of the greatest, or only importance, and they do not attempt to reflect or learn from experience (Powell, 1989). Despite this, reflection is an ability that can be taught (Boenink *et al.*, 2004; Wald *et al.*, 2009) yet it requires training, development and regular practice (Driessen *et al.*, 2005; Nicol and Macfarlane-Dick, 2006; Sandars, 2009; Oosterbaan *et al.*, 2010; Aronson, 2011; Aronson *et al.*, 2012).

One approach that has shown to support the development of individual reflective skills is regular peer-evaluation and face-to-face peer-feedback (Pee *et al.*, 2002; Wallman *et al.*, 2008; Sandars, 2009; McDonald, 2010; Subramanian *et al.*, 2013), with clear goal settings (Kamp *et al.*, 2013). If the process of reflection through peer-assessment and feedback can help to reframe and organise theory, then reflection becomes a vehicle for effective learning (Kaufman and Mann, 2010).

1.4.11 Clinical workplace-based assessment and the role of peers

The assessment of clinical competence is a critical issue as it relates directly to the quality of patient care (Nulty *et al.*, 2010; Albino *et al.*, 2012). The introduction and later development of the Objective Structured Clinical Examination (OSCE) (Harden and Gleeson, 1979; Schoonheim-Klein *et al.*, 2008; Schoonheim-Klein *et al.*, 2009), to assess students at the “Shows how” dimension of Miller’s model (1990) (graphically represented in Figure 1-5 in page 46), after a near exclusive reliance on using written and oral examinations with subjective standard settings and poor reliability (Norcini and McKinley, 2007), contributed to a significant improvement in the reliability of authentic *competence* tasks assessment (Norcini, 2005; van der Vleuten and Schuwirth, 2005; Norcini *et al.*, 2011).

However, what students “show how” to do in controlled high stakes examinations like the OSCE (showing competence), does not assure they will perform competently in actual practice (Rethans *et al.*, 2002). As a result, a challenging step has now been taken

to move the assessment venue from the relatively homogeneous setting of structured examinations, back to the uncontrolled and heterogeneous real world, assembling “*teaching, learning, and assessment*” in one single place, that is, the workplace (Norcini and Burch, 2007).

Though more research is needed (Kogan *et al.*, 2009), this process has been encouraged by studies showing good validity and reliability for some newly developed, less standardised methods of direct observation workplace-based assessment (WPBA) that focus on the “Does” top end of Miller’s pyramid (Miller, 1990), while trainees are in the real clinical setting of patient care (Norcini and McKinley, 2007).

Thus, these new assessment methods are the outcome of a thoughtful attempt during the last two decades to design structured and standardised assessment forms (questionnaires) of the quality of observed students’ behaviours (Cantillon and Wood, 2010) but in a naturalistic setting (Boursicot *et al.*, 2011), with the advantage of allowing for immediate formative feedback (Norcini and Burch, 2007). The most commonly found WPBA tools in the literature, and therefore subsequently described, are the Mini-Clinical Evaluation Exercise (mini-CEX), Direct Observations of Procedural Skill (DOPS), Longitudinal Evaluation of Performance (LEP), Case-base Discussion (CbD), Multi-Source Feedback (MSF) which includes the Patient Assessment Questionnaire (PAQ) and the Mini-Peer Assessment Tool (mini-PAT).

Mini-Clinical Evaluation Exercise (mini-CEX)

The mini-CEX, designed by the American Board of Internal Medicine, is used to assess trainees while conducting a clinical consultation (Norcini *et al.*, 1995), generally of an inpatient, outpatient or emergency department settings (Norcini *et al.*, 2003). Thus, the tutor observes the learner while interviewing a patient and/or conducting a physical examination, and judges this and her or his professionalism, clinical judgment, counselling, communication, organisation and efficiency, as well as the overall clinical competence. The trainee then summarises the encounter to the tutor by providing the diagnosis and/or treatment plan (Cantillon and Wood, 2010).

The encounter is recorded and scored on the mini-CEX form (Appendix 5) by the trainer, who uses this information to provide the trainee with structured feedback during debriefing (Cantillon and Wood, 2010). The encounter takes about 10 or 15 minutes with 5 minutes for feedback. Typically, the trainee would be assessed on several

occasions by different trainers, so that bias from a single assessor is reduced. Further, multiple mini-CEX on different patients addresses the case specificity of clinical performance (Norcini, 2011).

The mini-CEX is intended to identify the few trainees whose performance is totally unsatisfactory and to provide the remainder with the opportunity for ongoing formative assessment and feedback (Norcini and Burch, 2007). It has also been designed to ensure that trainees' clinical skills have been observed and assessed by staff members (Norcini, 2011).

There are copious studies on the use and characteristics of the Mini-CEX mainly in medicine (Norcini *et al.*, 1995; Norcini *et al.*, 1997; Norcini *et al.*, 2003; Hill *et al.*, 2009; Cook *et al.*, 2010; Dewi and Achmad, 2010; Pelgrim *et al.*, 2012b; Al Ansari *et al.*, 2013; Alves de Lima *et al.*, 2013; Liao *et al.*, 2013; Weston and Smith, 2014), but a few in dentistry as well (Millett, 2011; Kalsi *et al.*, 2013).

Direct Observations of Procedural Skill (DOPS)

The DOPS is a variation of the mini-CEX, designed by the Royal College of Physicians to assess and provide feedback on practical clinical procedures (Wragg *et al.*, 2003). Just as with the mini-CEX, trainees are observed and assessed by a staff member while practicing a clinical procedure on real patients. The scoring and the provision of feedback are similar to that of the mini-CEX. However, DOPS encounters are usually longer than those of the mini-CEX as they require the duration of the procedure being performed (Wilkinson *et al.*, 2008). Thus, DOPS assessment will normally review the indications for the procedure being performed, how consent was obtained, whether appropriate analgesia was used (if needed), technical ability to perform the procedure, professionalism, asepsis, awareness and management of complications, as well as the overall competence (Appendix 6) (Cantillon and Wood, 2010). Although the original DOPS format was designed as a generic tool that could be used in any clinical procedure (Wragg *et al.*, 2003), there are today large number of DOPS formats to assess different specialties both in medicine and dentistry (The Royal College of Surgeons of England, 2014).

Though not as many as with the mini-CEX, there are studies reporting the use of DOPS in medicine (Wilkinson *et al.*, 2008; Davies *et al.*, 2009; Ahmed *et al.*, 2011; Mitchell *et al.*, 2011; Bindal *et al.*, 2013; Cobb *et al.*, 2013; Delfino *et al.*, 2013; Watson *et al.*,

2014). However, and although Kalsi *et al.* (2013) included DOPS in their review of several WPBA methods for foundation and postgraduate dental training, and that the UK Intercollegiate Surgical Curriculum Programme (ISCP) presents twelve different specialty DOPS (The Royal College of Surgeons of England, 2014), there are no published studies in dentistry reporting its use. It can be speculated that this is probably due to the series of published reports directed by Linda Prescott-Clements on competency-based assessment of postgraduate dental training in Scotland (Prescott *et al.*, 2002). Thus, she developed the Longitudinal Evaluation of Performance (LEP) and published it before the appearance of DOPS.

Longitudinal Evaluation of Performance (LEP)

As mentioned above, Prescott-Clements *et al.* (2002) introduced a method of clinical assessment, the Longitudinal Evaluation of Performance, which considered, as identified by the authors, three major challenges that clinical educators face when designing competency-based systems of assessment. These were *i)* the need to assess different areas of clinical competence; *ii)* the necessity of alignment between assessment and training objectives, and *iii)* the type and focus of the assessment method.

Consequently, as it can be appreciated from Appendix 7, the LEP eight domains look like a combination of the mini-CEX and DOPS, as, besides generic skills like “professionalism” and “communication”, it includes for example “consultation skills” and “clinical judgement and diagnosis” typically from the mini-CEX, but also “technical ability and manual dexterity” more distinctive of DOPS. A latter version of the same instrument replaced the original last question “Overall competence” for the “Trainee’s insight into performance” in order for trainees to develop this competence with experience and regular feedback. The satisfactory implementation of this improved version of the LEP tool in postgraduate dental trainees in Scotland has been recently published by Prescott-Clements *et al.* (2011). Thus, the LEP has been reported to combine a strong formative approach through continuous assessment and systematic feedback, which, within a wider assessment programme, contributes towards a summative decision of competence (Prescott-Clements *et al.*, 2008).

A slightly modified version of the LEP, named “Dental Evaluation of Performance” (DEP), has been reported to be part of the workplace based assessment tools for the dental foundation training in the English Mersey deanery (Grieveson *et al.*, 2011a), and more

recently in other 12 deaneries in England and Northern Ireland (Kirton *et al.*, 2013).

Case-base Discussion (CbD)

The CbD is the UK variation of the Chart-Stimulated Recall (CSR) developed in the United States to be used in the context of emergency medicine (Maatsch *et al.*, 1983). In both cases, the trainees chose several patient records whom they have recently seen and in which they have made notes (Cantillon and Wood, 2010). Subsequently, in an encounter of approximately 20 minutes (Norcini and Burch, 2007), the assessor selects one record for discussion and asks the trainee to describe the case and asks clarifying questions directed to know why the trainee acted as she or he did (Norcini, 2011). Upon relevant details of the case emerge, the assessor focuses the discussion on the trainees' diagnostic reasoning, investigation, the rationale for choosing a certain action and their awareness of differential diagnoses, ethical issues, follow-up and future planning. Further, since patient records are available at the time of the encounter, record keeping can also be assessed as the CbD forms usually contain a domain for this purpose (Appendix 8) (Norcini and Burch, 2007).

It is expected that trainees participate in multiple encounters so that several clinical cases are discussed with different trainers (Setna *et al.*, 2010) in order to reduce any bias caused by case specificity (Cantillon and Wood, 2010).

Extensive studies have reported the use of CbD in foundation medical doctors (Mitchell *et al.*, 2011; Mitchell *et al.*, 2013), as well as in medical specialty trainees (Setna *et al.*, 2010; Sabey and Harris, 2011; Mehta *et al.*, 2013). Further, the use of CbD has also been reported in dental foundation training (Grievesson *et al.*, 2011a; Kirton *et al.*, 2013).

Multi-Source Feedback (MSF)

The idea behind MSF is to directly assess "routine practice" which has been described to be much harder to assess compared to single encounter assessment (Cantillon and Wood, 2010), such as the ones previously described. Thus, MSF, also known as 360-degree assessment, represent the systematic collection of colleagues' and patients' perspectives in order to assess performance and to provide feedback to students and trainees. In the case of patients, they are usually asked to rate the trainee's communication skills and professionalism (Goldie, 2013) using specifically developed forms (Appendix 9) (Hurst *et al.*, 2004).

A number of different tools regarding peer-assessment have been reported. Evans *et al.* (2007) used the Objective Structured Assessment of Technical Skills (OSATS) scale, developed to assess technical skill of surgical trainees, to compare single occasion peer-assessment scores of 38 postgraduate trainees, specialising in oral and maxillofacial surgery, to those assigned by trainers, following a third molar surgery.

Lurie *et al.* (2006) presented an online version of the Peer-Assessment Protocol (PAP), originally developed by Dannefer *et al.* (2005), to assess behavioural professional competencies like “work habits” and “interpersonal sensitivity” of undergraduate medical students. Recently, Mackillop *et al.* (2011a) reported the feasibility of generic on-line Multi-source feedback questionnaire for revalidation of UK career-grade doctors.

Archer *et al.* (2005; 2010) presented the 24-item Sheffield Peer Review Assessment Tool (SPRAT), which reliably, yet feasibly allowed senior medical paediatricians in training to assess each other’s clinical competencies. The same group (Archer *et al.*, 2008) reported later the validation of a simplified SPRAT instrument, namely the 16-item Mini-Peer Assessment Tool (mini-PAT) (Appendix 10).

In the mini-PAT, peers’ assessment is given anonymously (Norcini and Burch, 2007). Typically the trainee selects eight assessors representing a mix of senior supervisors, trainee and nursing colleagues, clinic staff and so on, who are all requested to complete the mini-PAT and to return them to a central location for processing. The trainee also self-assesses, completing the same form. Subsequently, all questionnaires data are amalgamated and presented to the trainee by a supervisor in a manner that she or he can compare the self-assessment with the mean ratings of the peer assessors. Trainee and tutor then agree on strengths and what aspects of clinical, professional or team performance need more work. It is normal practice in the UK Medical Foundation Programme to perform this assessment twice a year during the duration of the training programme (Norcini, 2011).

The mini-PAT has been used in the medical foundation assessment programme in conjunction with some of the previously described WPBA forms (Archer *et al.*, 2008; Davies *et al.*, 2009; Mitchell *et al.*, 2011; Mitchell *et al.*, 2013), in medical surgical trainees (Pereira and Dean, 2009; Eardley *et al.*, 2013), and in postgraduate pharmacists (Davies *et al.*, 2013). In dental education, the use of the mini-PAT has been suggested (Mattheos *et al.*, 2009) but there are no publications reporting its use. However, as part

of a multisource assessment the Patient Assessment Questionnaire (PAQ) has been used for Foundation Dental Training in the UK (Grievesson *et al.*, 2011a; Kirton *et al.*, 2013).

Studies reviewing these and other WPBA methods have shown positive and sufficiently reliable and valid results to be used for formative purposes (van der Vleuten and Schuwirth, 2005; Norcini and Burch, 2007; Cohen *et al.*, 2009; Davies *et al.*, 2009; Kogan *et al.*, 2009; Boursicot *et al.*, 2011; Norcini *et al.*, 2011; Pelgrim *et al.*, 2011; Driessen *et al.*, 2012; Donnon *et al.*, 2014).

However, a number of potential concerns over the current practice of WPBA have also emerged (Spencer, 2003; Pereira and Dean, 2009; Saedon *et al.*, 2010; Archer and McAvoy, 2011; Beard, 2011; Al-Kadri *et al.*, 2013; Driessen and Scheele, 2013; Ferguson *et al.*, 2014). A review paper by Miller and Archer (2010), pointed out that very few published articles explore the impact of WPBA on trainees' education and performance. They showed that MSF can lead to performance improvement while there was no evidence that DOPS, Mini-CEX or CbD did so. In a later study of 52 general medical practitioner specialist trainees conducted by Sabey and Harris (2011), 74% reported Multi-Source Feedback to be the most useful tool, whilst 35% did for CbD, 14% for DOPS and only 10% for Mini-CEX. While 55% found WPBA useful as a learning tool, only 45% reported the use of WPBA for summative assessment to be helpful in identifying a struggling doctor. Trainees place a low value on rating scale scores and they perceive a lack of honesty in assessments, as well as bias and a "box ticking" attitude that undermines the credibility of WPBA. In another recent study (Quantrill and Tun, 2012), 95% of trainees reported problems in finding an assessor to complete their WPBA. Although feedback should be at the very heart of WPBA, only 43.9% of trainees received immediate feedback and 22% received it one week later. The reason for this lack of feedback has been attributed to senior trainees or clinical supervisors with limited time or incentive to give due attention to WPBA (Quantrill and Tun, 2012). Thus, Saedon *et al.* (2010) emphasised the need to train staff on the great importance of quality feedback (including theory), and trainees around the need for feedback, how to give and receive it. Similarly, Bindal *et al.* (2011) have called for additional training especially on the value of WPBA and consultants to have protected time in their job plans for training.

Additionally, Norcini and Burch (2007) reported the apparent poor Faculty participation

in formative assessment and feedback strategies as probably the most significant limiting factor currently identified in WPBA. One challenge set by the authors is to make current WPBA methods more user-friendly by their modification and simplification, especially as they are designed to be used in busy clinical settings.

Fortunately, published studies using WPBA in dentistry have shown somehow better outcomes. Thus, Prescott-Clements *et al.* (2008) reported positive results to support the validity, educational impact and feasibility of the LEP, using almost 10,000 completed forms submitted over 2 years by 201 postgraduate trainees. Later, using the subsequent version of the LEP, previously described, the same researchers (Prescott-Clements *et al.*, 2011) once again obtained encouraging results when tutors assessed and provided feedback to 139 postgraduate dental trainees to develop their insight into their performance.

In another study of 41 Foundation dental practitioners, Grieveson *et al.* (2011a) surveyed their perceptions of the effectiveness of three WPBA methods used during their training: the Dental Evaluation of Performance (D-EP), Case-based Discussion (CbD) and Patient Assessment Questionnaire (PAQ). Overall 84.1% of trainees felt that WPBA helped them improve patient care (96% with CbD). More specifically, 81.6% of trainees felt that feedback provided after being assessed by tutors with DEP highlighted areas to develop their learning better than when CbD (76.8%) or PAQ (60.9%) were used. Further, a large majority (>79%) found that feedback gave them an insight into their own development needs.

A similar but larger study of 359 Dental Foundation trainees published later by the same group (Kirton *et al.*, 2013), showed that the overall experience of WPBA was positive and played an important role in trainees' learning and building confidence. Interestingly, in a further question not included in the initial report, the majority of trainees (>62%) felt that feedback received from tutors enabled them to be reflective in their clinical practice. Further, trainers and trainees found CbD the most beneficial of all the WPBA used tools. However, in agreement with reports from medical education (Saedon *et al.*, 2010), the authors highlighted the need for comprehensive training in the WPBA tools to ensure their efficacy.

Despite the widespread dissemination of WPBA practices in medical as well as in dental foundation programme and postgraduate training, there is a lack of literature of its use in undergraduate medical and dental education. Among the very few studies

available, Nesbitt *et al.* (2013) surveyed 288 year-4 medical students in two consecutive cohorts, to collect their experiences on the use of WPBA. A total of 46% (133) of the students found that feedback from the WPBA was useful (19% disagree or strongly disagree). An interesting question assessed students' perception on whether WPBS were a useful way of making sure that supervising doctors spent time with them while with patients and discussing cases: 42% agree or strongly agree, while 36% disagree or strongly disagree.

More in the psychometrics, Hill *et al.* (2009) estimated the validity and reliability of a modified mini-CEX when staff assessed medical undergraduate students. The authors found good overall utility of the modified mini-CEX for assessing aspects of the clinical encounter. Strengths included fidelity, wide sampling, perceived validity, and formative observation and feedback. Reliability assessed by means of Cronbach alpha was α 0.73 for 15 encounters.

While the mini-CEX has been described as a single WPBA encounter tool (Cantillon and Wood, 2010), Playford *et al.* (2013) used it as a staff performed longitudinal assessment of medical undergraduate students over a three-year period. They analysed 5.686 mini-CEX completed forms and found a Cronbach alpha reliability of $\alpha=0.80$. Marks were significantly affected by the grade of the marking tutor, difficulty of the clinical encounter, and the clinical discipline. An increase in mini-CEX marks over the course of the academic year was also noticed, especially during the initial formative months. The authors concluded that the longitudinal assessment identified and contributed to students' skills development. Further, students actively participated in their own development by using the mini-CEX in excess of course requirements and seeking harder markers.

Finally and closer to the aim of the present study, Bennett *et al.* (2012) piloted the utility and students' acceptability of the mini-CEX as a framework for peer-feedback. Thus, 40 undergraduate medical students, working in pairs (alternating learner and assessor roles), undertook two peer mini-CEX evaluations whilst on clinical attachments. Accordingly, they marked and identified areas for improvement in their peer's performance. The results showed that students liked to compare their performance, being at the side of the examiner, and receive *useful* peer-feedback. However, they questioned the validity of the peer-feedback, and asked for more guidance regarding the standards for peer-assessment. The authors reported the mini-CEX to be a useful

framework for formative peer-feedback of undergraduate students. Further, peers were able to identify those areas of good performance as well as those where improvement was required. However, in contrast to the provision of useful feedback, peers did not score less than 4 out of 5, which, according to the authors, was unlikely to reflect the real range of the students.

1.5 Literature review conclusions

The literature on education is complex, diverse, rich in content and far more extensive than can all be included in these pages. At the same time, many opinions, principles and theories are argued and contra-argued constantly which enriches it even more. This has allowed the accumulation of a huge body of educational theory in the social sciences that could be used to inform daily theoretical as well as clinical teaching practice. However, the present review seems to show that the relationship between this rich accrued knowledge and healthcare education could be improved for the benefit of our teaching. It appears that healthcare academics, who share and contribute to this educational knowledge, struggle to demonstrate to fellow practitioners the usefulness of the guiding principles it can provide to inform their teaching practice (Kneebone, 2002). Further, educational textbooks are, arguably, often associated with laboratory experiments and thus, far from the reality of classroom practice (Harden and Laidlaw, 2013). Perhaps debatably, medical education feels privileged and sets itself apart from the monotony of higher education (Swanwick, 2011), as if it had nothing to learn from other fields of education.

The strong evidence supporting the benefits of feedback with very specific characteristics, as well as the implementation of peer-assessment practices focused on its learning potential, not just as a marking exercise, that are largely available in the educational literature, would lead one to expect this to be reflected in the dental education literature. As shown above, this is not the case. However, this does not necessarily mean these practices are not being implemented in dental education, but if we only take the example of the NSS on dental students' feedback perception compared to that from education students (Table 1-2 in page 64), it appears as if the evidence is not being used to its full potential.

The wide theoretical and practical background of peer-assessment from other areas of education, as presented above, provides support and underpins the further study of peer-

assessment and peer-feedback in dental students as a possible learning experience. The high contact time dental clinical partners share in their respective workplace, while practicing their skills or during patient treatment makes it an ideal environment for them to help each other through formative peer-assessment and peer-feedback, provided there has been careful and rigorous planning, including student training.

The same literature highlights the challenges dental students' peer-assessment experience might encounter, among these are reliability and validity issues, its formative or summative purpose, students' worries and confidence issues, acceptability and level of participation, domains to be assessed, and so forth.

Finally, on implementing this knowledge, and taking Watling *et al.* (2013) and their study on medical and music learners as a model, we would be pleased to see our students aim for an ever-better performance, learn teaching and self-assessment skills, but recognise and seek life-long external feedback.

Chapter 2

Teaching and Assessment Approaches in Dental Education

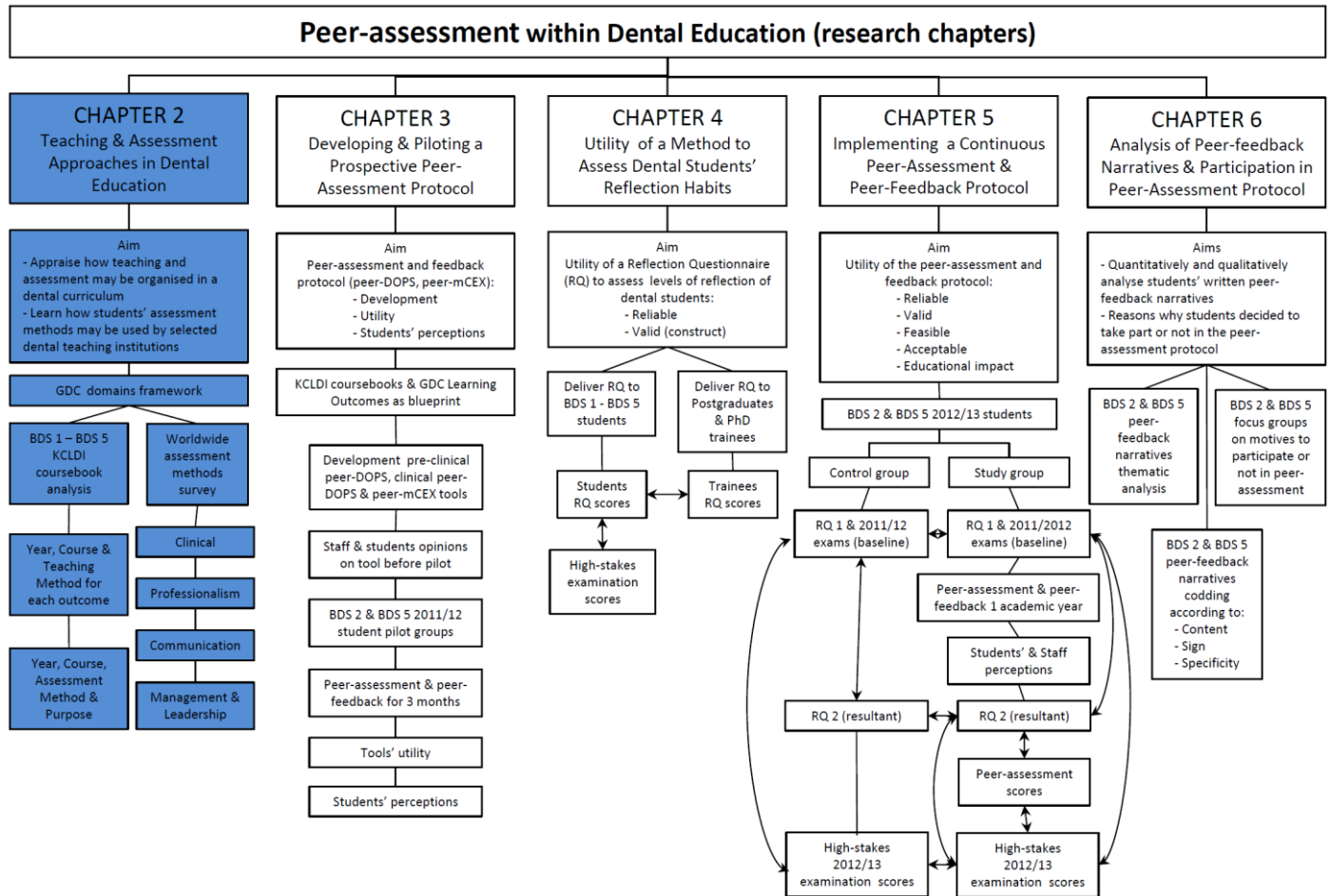


Figure 2-1 Flowchart of the study's research chapters. The second current chapter is highlighted to help in providing an overall view.

2.1 Introduction

Dental education requires the integration and broad understanding of healthcare and basic and oral sciences for students to develop their knowledge, motor and intellectual skills, as well as attitudes and behaviours (Entwistle and Hounsell, 1977; Oliver *et al.*, 2008). For dental education to be successful, particularly in relation to developing habits that encourage a self-directed learning culture (Manogue *et al.*, 2011), it requires a wide range of teaching methodologies as well as an integrated assessment strategy.

Dental schools in the United Kingdom comply with learning outcomes determined by the General Dental Council (GDC) as the regulatory body. The GDC was created in 1956 (Basker, 2006) to ensure high standards of dental education (UK Government

"Dentist Act 1984") at all its stages (General Dental Council, 1997). The initial GDC "Recommendations concerning the Dental Curriculum" (1990) evolved into a sequence of three publications entitled "The First Five Years" (1997; 2002; 2008) with increasing detail regarding the expected competency profile of a new graduate. In January 2012 the GDC published its latest guidelines for course providers and awarding bodies entitled "Preparing for Practice: Dental team learning outcomes for registration" (2012a). This document sets out the current learning outcomes, grouped into four integrated and mutually supported domains (Figure 2-2), that a candidate must be able to demonstrate by the time of graduation.

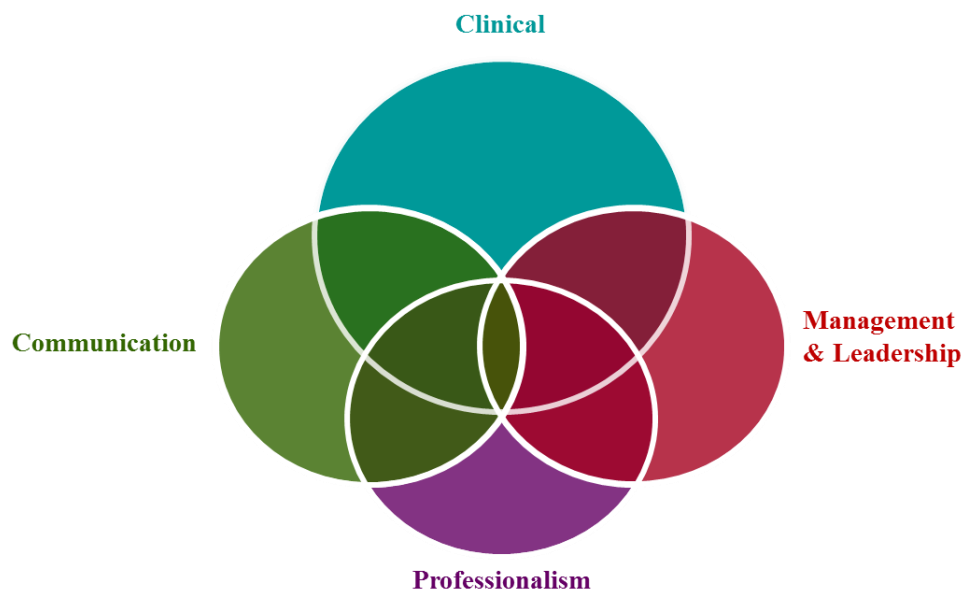


Figure 2-2 "Preparing for Practice: Dental team learning outcomes for registration" four integrated and mutually supported domains (2012a p. 9).

These learning outcomes reflect the already required "knowledge, skills and attitudes" outlined in the two last versions of The First Five Years (2002; 2008), with the addition of "behaviours" with the aim of developing a rounded professional who practises safely, effectively and professionally (General Dental Council, 2012a). Whilst the 2002 and 2008 versions identified three levels to a developmental framework for each learning outcome ("be competent at", "have knowledge of" and "be familiar with"), these have now been replaced by a different analytical framework (Pangaro and ten Cate, 2013). This consists of a list of verbs called "keywords" associated with the learning outcomes required levels of Knowledge, Skills and Attitudes/Behaviours (Table 2-1).

Table 2-1 Taxonomy used to approach the new 2012 GDC Learning Outcomes (General Dental Council, 2012a).

Keywords used in the learning outcomes		Application
Knowledge	Describe, recognise, explain, discuss, interpret, identify, evaluate	Recall or recognise information, explain or interpret meaning from a given scenario or statement
Skill	Use, apply, manage, produce, implement, perform, record, extract, modify, refer	Use or apply knowledge and skills
Attitudes / Behaviours	Participate, contribute, act, take responsibility, respect	Receive and respond to information, react and participate actively, prioritise and display values

Regarding assessment, the GDC “Preparing for Practice” (2012a) document states that students should be assessed rigorously, appropriately and reliably in a demonstrable and clear way against all outcomes. However, it only mentions some assessment methods such as workplace based assessment, portfolios, projects and examinations in passing. Later that year the GDC published the “GDC Standards for Education” in which the third of its four standards concentrated on “Student Assessment” (General Dental Council, 2012c pp. 6-8). Whilst it does not restrict or impose any assessment tool, leaving this to each institution to decide, it does mention the need for ‘reliable’ and ‘valid’ assessment methods to demonstrate achievement of the GDC learning outcomes. Indeed, it sets two important and useful requirements relevant to this study:

- Requirement N. 20: “providers should seek to improve student performance by encouraging *reflection* and by providing *feedback*” which, among other proofs, means the need for training in reflection and receiving feedback, and some evidence of reflection.
- Requirement N. 25: asks for valid and reliable “multiple samples of performance” and evidence of “continuous assessment” (General Dental Council, 2012c).

Finally, the GDC regulatory document entitled “Quality Assurance Process” (General Dental Council, 2012b) requires from every new programme a mapping of assessment against the “Preparing for Practice” learning outcomes.

Amalgamating these together, there is a clear message from the regulator to dental educational providers to collect evidence and assume the role and responsibility of preparing students to carry out reflective practice and self-directed learning, to adhere to

a life-long learning (General Dental Council, 2012a).

It is not surprising that King's College London "Learning, Teaching and Assessment Strategy" (King's College London Academic Services, 2011) states that students should become increasingly creative, critical and analytical and be able to work independently, taking their own decisions with a sound base of values and an understanding of others. The King's Strategy goes on to assert that assessment is an integral part of the learning process and must be fair, transparent and encourage learning. Hence, by engaging students in the assessment process and offering them feedback in a notion of "assessment *for* learning" (Schuwirth *et al.*, 2011), it is designed to enable students to continually improve their work.

In view of the above we felt it appropriate to investigate how dental teaching and particularly assessment are conducted using the GDC learning outcomes as the framework.

2.2 Aim

To investigate the different teaching and assessment methodologies used in dental education using King's College London Dental Institute (KCLDI) as true model.

To investigate assessment practices and strategies used by selected dental teaching institutions around the world to measure students' progress particularly in relation to the use of peer-assessment.

2.3 Hypotheses

- i. KCLDI encompasses the teaching and assessment of knowledge, skills and attitudes/behaviours in all four GDC domains (Clinical, Communication, Professionalism, and Management and Leadership).
- ii. Peer-assessment is employed in more than 50% of the surveyed (inter)national dental teaching institutions.

2.4 Materials and Methods

2.4.1 KCLDI teaching and assessment

All 51 undergraduate coursebooks (syllabus) that constitute the Bachelor of Dental Surgery (BDS) 2011/2012 program at KCLDI were collated. A descriptive quantitative analysis of these was undertaken to determine the current teaching and assessment practices.

The “Preparing for Practice” (General Dental Council, 2012a) learning outcomes for dentists were then used as a framework to map every teaching and assessment session from the KCLDI syllabus using the template shown in Table 2-2. The teaching methodologies at each BDS year were then classified in six clusters: Lectures, Tutorials & Workshops, Self-directed Learning, Practicals & Clinical, E-learning and others. Similarly, assessment sessions were grouped by year and then categorized according to Miller’s Pyramid of Professional Competence (Miller, 1990).

Table 2-2 Template used for mapping the 2012 GDC learning outcomes with BDS teaching and assessing methodologies at KCLDI.

GDC		Teaching			Assessment		
DOMAIN	OUTCOMES	Year/module in which the learning outcome is covered	Course in which the learning outcome is covered	How the learning outcome is covered	Year assessed	Course learning outcome is assessed	How the learning outcome is assessed
Clinical							
Communication							
Professionalism							
Management and Leadership							

The KCLDI BDS programme learning outcomes were categorised according to the new GDC taxonomy of “knowledge”, “skills” and “attitudes/behaviours” keyword verbs (Table 2-1) for each of the four domains (Clinical, Communication, Professionalism, and Management and Leadership).

2.4.2 Assessment practices and strategies online survey

The resulting list of KCLDI assessment methodologies used across the BDS programme were added with other tools from the dental (Manogue *et al.*, 2001; Albino *et al.*, 2008;

Kramer *et al.*, 2009) and medical (Wass *et al.*, 2001; Shumway and Harden, 2003; Norcini and Burch, 2007; Kogan *et al.*, 2009; Swanwick, 2011) education literature to prepare an assessment practices online survey. These assessment tools list were then mapped to the four GDC “Preparing for Practice” domains (Clinical, Communication, Professionalism, Management and Leadership). To enhance the survey content validity (Streiner and Norman, 2008), the resultant mapping was reviewed by seven internationally renowned dental educators (all members of the Association for Dental Education in Europe (ADEE)), at the “Assessment of Clinical Competencies in Dental Education” Special Interest Group (SIG).

The outcome of this review was presented as an English language online survey (Table 2-3) and distributed online using Survey Monkey® (Palo Alto, California, USA) to selected dental teaching institutions to assess their assessment strategies for measuring students’ progress within each of the four GDC domains. The selected institutions were Dental Schools, Institutes or Centres involved in research publications in the European Journal of Dental Education and/or the Journal of Dental Education between the years 2009 to 2012 with an English website¹.

In total, 49 dental teaching institutions, representing 15 countries from all 4 continents were selected and their Dean or Director of Education sent an electronic invitation explaining the aims of the study inviting them to consent and complete the online survey. They were asked to identify which of the assessment methods in the list were the most important sources of information to routinely assess students' competence and performance (not seldom-used techniques) at their Dental School / Institute / Centre. They were also requested to choose whether the assessment method was used formatively, summatively or as a hurdle (gateway).

All collected data were entered into an Excel (Microsoft Inc., Seattle, WA, USA) spreadsheet and descriptively analysed using the Statistical Package for the Social Sciences® Windows® version 21 (SPSS Inc. IBM, Chicago, IL, USA).

¹ USA dental teaching institutions were excluded as they had recently been surveyed in a similar study Albino, J., Young, S., Neumann, L., Kramer, G., Andrieu, S., Henson, L., Horn, B. & Hendricson, W. 2008. Assessing Dental Students' Competence: Best Practice Recommendations in the Performance Assessment Literature and Investigation of Current Practices in Predoctoral Dental Education. *Journal of Dental Education*, 72, 1405-1435.

Table 2-3 Survey designed to assess how dental teaching institutions assessed each of the four GDC “Preparing for Practice” domains.

Assessment method	Domains			
	Clinical	Communication	Professionalism	Management and Leadership
• Multiple choice questions <u>linked to cases</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Multiple choice questions <u>not linked to cases</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Short Answer Questions (complete the sentence), quiz, brief tests and questionnaires	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Essays	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Oral examinations <u>with checklist</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Oral examinations <u>without checklist</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Research projects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Poster or other presentations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Observation of students' performance of technical skills in preclinical labs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Observation of students' performance in computer-based or other types of simulations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Chart-stimulated oral exams (clinical reasoning examinations)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Objective Structured Clinical Examinations (OSCE)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Structured Clinical Operative Test (SCOT)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Triple Jump (1st: Patient interview & examination observation, 2nd: Write assessment and treatment plan, 3rd: Oral exam on examined patient)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Structured Workplace-Based Assessment of students' performance (WPBA) <u>termly</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Structured Workplace-Based Assessment of students' performance (WPBA) <u>every patient</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Structured Workplace-Based Assessment of students' performance (WPBA) <u>occasionally</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Students self-assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Students peer-assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Assessment by patients	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Portfolios	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Learning diaries, log books, record review	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Quantity / Number of procedural units performed by students	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Other Assessment Methods (Please specify which ones)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2.5 Results

2.5.1 KCLDI teaching and assessment

The complete transcription of the KCLDI syllabus mapped to the GDC learning outcomes framework is included in the Appendices (Appendix 11).

Table 2-4 KCLDI teaching methodologies by course year.

Teaching Methodologies (sessions)							
Course Year	Lectures	Tutorials & Workshops	Self-directed learning	Practicals & Clinical	e-learning	Others *	Total
BDS 1	188	65	9	35	7	11	315
BDS 2	166	89	58	120	1	4	438
BDS 3	147	9	15	153	48	1	373
BDS 4	114	20	12	150	10	2	308
BDS 5	65	16	0	193	6	2	282
Total	680	199	94	651	72	20	1716

* poster, miscellaneous

The total number of teaching sessions for the five years of the KCLDI BDS programme was 1716 (Table 2-4). The commonest teaching methodology was Lectures (40%), followed by Practical & Clinical (38%), Tutorials & Workshops (12%), Self-directed learning (5%), e-learning (4%) and others (poster, miscellaneous) (1%) (Figure 2-3).

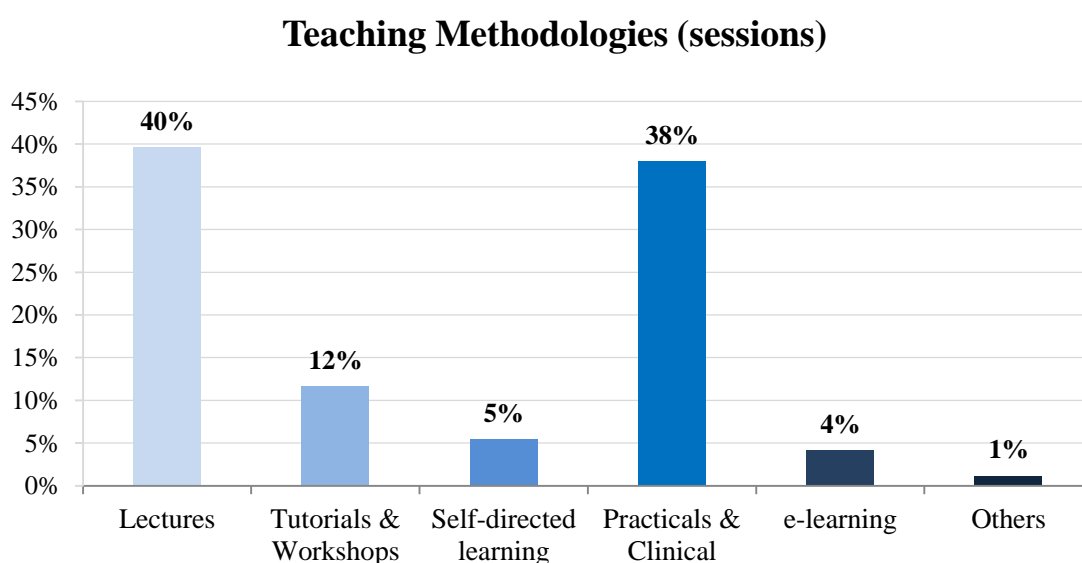


Figure 2-3 Graphical representation of the percentage of teaching methodologies from all courses together.

With respect to assessment, the five year programme included 23 different assessment methodologies totalling 153 assessment sessions including formative (55%), summative (30%) and hurdle (15%). The most frequently used assessment tools were structured essays (13%), workplace based assessments marked by tutors (13%) (Tutor Mark Assessments (TMA), Structured Clinical Operative Tests (SCOT), Student Performance Indicators, Paediatric Clinical Competency), Single Best Answers (12%), Extended Matching Questions (11%), OSCEs (9%), quiz, brief tests and questionnaires (10%), Short Answer Questions (7%), oral examinations (5%), skills laboratory practicals (5%), case presentations (5%), portfolios (5%), self-assessment (3%), Patient Assessment Questionnaires (1%), and clinical reasoning examinations (1%).

Remarkably, these assessment methods took different roles as formative, summative or hurdle. Thereby, from the total of 153 assessment sessions, those with a formative purpose were represented mostly by quiz, brief tests and questionnaires, Single Best Answers, workplace-based assessments, Extended Matching Questions, OSCE, and Portfolios. Differently, summative assessment methods included mainly Essays, Short Answer Questions, Single Best Answers, and Extended Matching Questions. Finally, Hurdle intentioned assessments included predominantly workplace-based Tutor Mark Assessments, Orals, and case presentations.

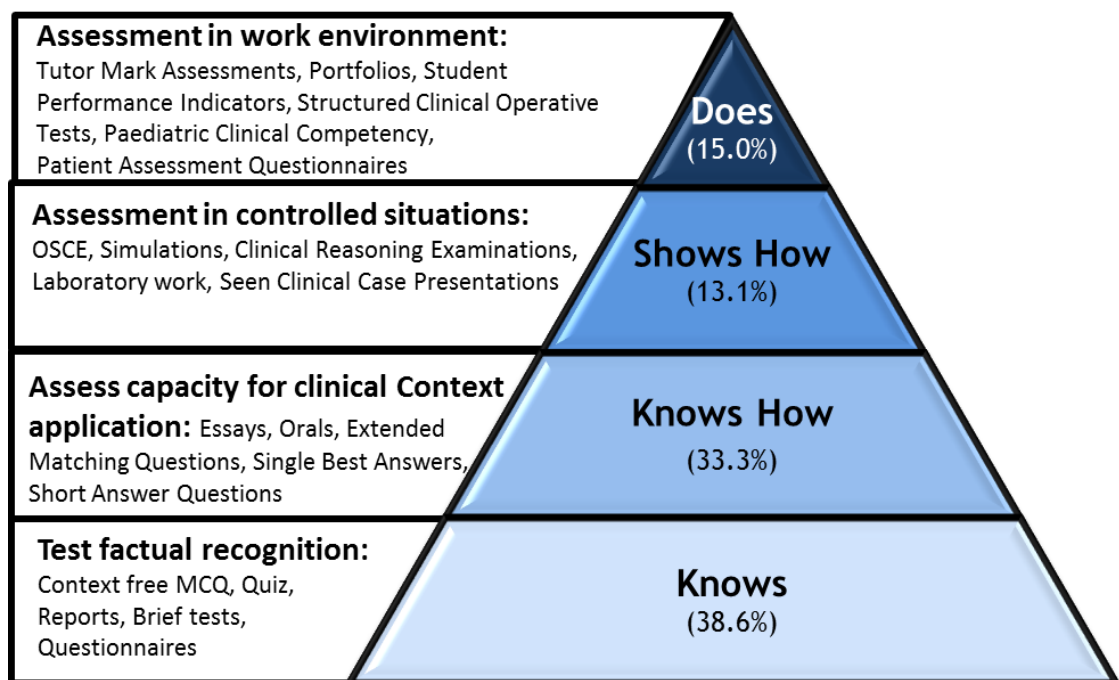


Figure 2-4 Overall assessment framework within the four levels of Miller's Pyramid (Miller, 1990).

When the assessment methods from the framework were classified using the four levels of Miller's Pyramid of Professional Competence (Miller, 1990), it was possible to appreciate (Figure 2-4) that the "Knows" level represented the predominant (38.6%) category followed by "Knows How" (33.3%), "Shows How" (13.1%) and "Does" (15.0%).

The overall picture however, changed dramatically when these assessment methodologies were divided according to their purpose (formative, summative or hurdle) (Figure 2-5). Particularly interesting is the low 1% of workplace assessment in relation to the summative group. This might denote the previously mentioned difficulties of clinical instructors to reliably assess learners' specific struggles (Audétat *et al.*, 2013) as well as clinical and reasoning skills, while at the same time ensuring high-quality patient care (Bowen, 2006; Norcini and Burch, 2007).

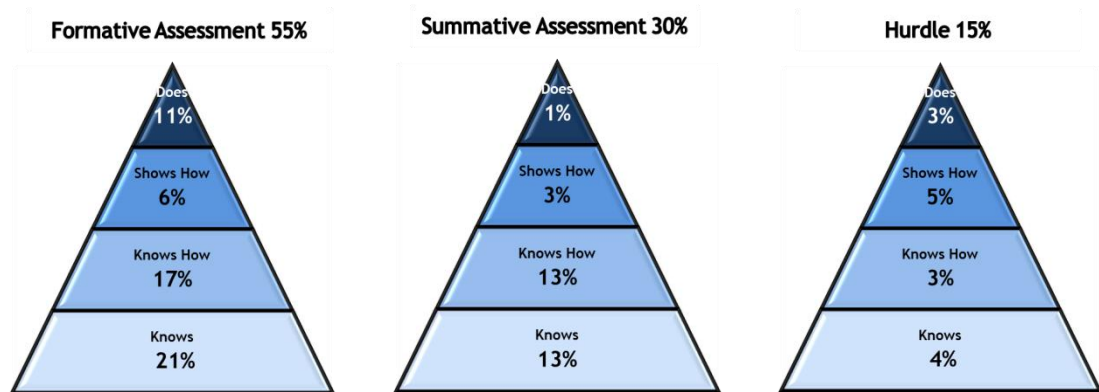


Figure 2-5 Summative, Formative and Hurdle assessment approaches mapped to the four levels of Miller's Pyramid (Miller, 1990).

Additionally, the distribution of these assessment types varied across the different BDS years (Figure 2-6). Accordingly, the test of factual recognition – Knows – through for example MCQs, together with the higher level assessment of clinical context applications – Knows How – through for example SBA and Essays, represented the majority of all assessment sessions throughout the programme. However, both these groups of assessment methods showed a clear decrease in their proportional representation from BDS 1 (45.5% for 'Knows' and 40.9% for 'Knows How') to BDS 5 (33.5% for 'Knows' and 28.6% for 'Knows How'). Interestingly, and though the fraction of assessment methods at the 'Shows How' level remained below the 17% all

over, it represented a vital stage for students to demonstrate their capacity to reflect and integrate all their professional skills through, for example the Clinical Reasoning Examination and the Clinical Case Presentation. This however paved the way for the higher ‘Does’ level of workplace assessment methods to overtake the ‘Shows How’ level already at BDS 3 (14.7% versus 11.8%, respectively), reaching a higher 21.9% and 21.4% at BDS 4 and BDS 5, respectively.

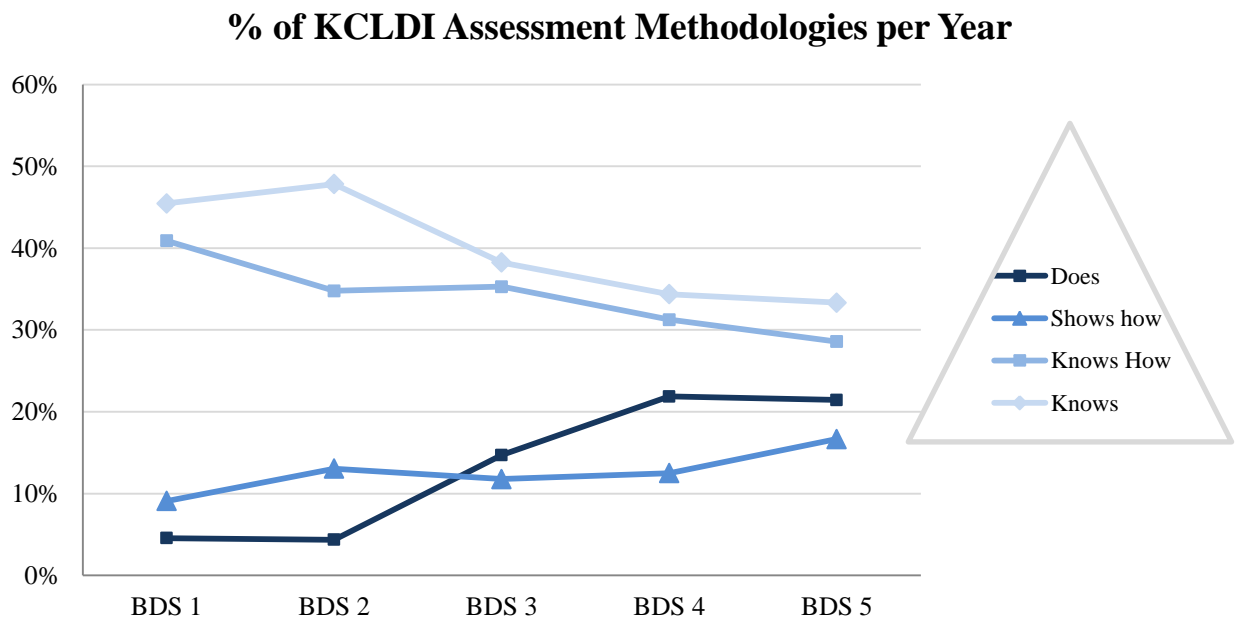


Figure 2-6 Percentage of assessment sessions grouped according to Miller's pyramid of professional competence, per year.

The “Preparing for Practice” (General Dental Council, 2012a) guidelines for dentists contain a total of 157 learning outcomes, of which 7 are overarching while 150 are domain specific (95 Clinical, 21 Communication, 19 Professionalism, and 22 Management and Leadership). The KCLDI BDS programme accounted for 142 (90%) of these learning outcomes, which means 15 were not formally covered: 5 in Clinical, 2 in Communication, 4 in Professionalism, and 4 in Management and Leadership domains (Appendix 11 Transcription of the KCLDI syllabus). However, the programme also included a number of 71 teaching sessions in the Clinical domain that did not match any of the 2012 “Preparing for Practice” learning outcomes (Appendix 11 Transcription of the KCLDI syllabus). One should keep in mind that the studied KCLDI 2011/2012 academic programme was developed to comply with the 2002 GDC guidelines (General Dental Council, 2002). It is expected that missing learning outcomes from the 2012 “Preparing for Practice” guidelines are to be gradually introduced in the subsequent

academic years and certainly with the new KCLDI curriculum introduced in 2013.

Coming back to the studied KCLDI BDS programme, the vast majority of the covered learning outcomes corresponded to the Clinical domain (67.4%), followed by Management and Leadership (13.7%), Professionalism (12.0%) and Communication (6.9%). According to the GDC taxonomy, 53.6% of these learning outcomes were classified in the “knowledge” category, whilst 32.2% corresponded to “skills” and 14.2% to “attitudes/behaviours”. These figures vary considerably when outcomes were analysed by domain (Table 2-5). Further, every single learning outcome was aligned with a teaching and assessment methodology as described in Appendix 11.

Table 2-5 Learning outcomes range according to the keyword verbs from the proposed taxonomy for each of the four domains.

Learning Outcome Taxonomy	KCLDI learning outcomes according to the GDC domains				
	Clinical	Communication	Professionalism	Management and Leadership	Total
	%	%	%	%	%
Knowledge	58.6	37.5	42.9	46.9	53.6
Skills	38.2	50.0	10.7	12.5	32.2
Attitudes / Behaviours	3.2	12.5	46.4	40.6	14.2
Total	67.4	6.9	12.0	13.7	100

2.5.2 Assessment practices and strategies online survey

The assessment survey was answered by 38/49 (78%) of the selected dental schools / institutes / centres within three weeks. The results (Table 2-6) demonstrate that all listed assessment methods were used by at least one surveyed institution, and reveals the wide variety of assessment strategies utilised to assess different domains.

The most widespread assessment method was the OSCE used for the majority (52.6%) of the institutions for the assessment of clinical and communication skills. Subsequently, other commonly utilised methods for the clinical domain were the MCQ linked to cases (39.5%), observation of students' performance of technical skills in preclinical labs (31.6%), structured workplace-based assessment of students' performance (WPBA) with every patient (31.6%), short answer questions (28.9%), and both essays and self-assessment which reported the same usage (26.3%).

Table 2-6 Survey results showing the % of dental teaching institutions using the listed assessment methods for each of the four GDC “Preparing for Practice” domains.

Assessment method	Domains			
	Clinical	Communication	Professionalism	Management and Leadership
• Multiple choice questions <u>linked to cases</u>	39.5%	15.8%	15.8%	13.2%
• Multiple choice questions <u>not linked to cases</u>	21.1%	2.6%	5.3%	7.9%
• Short Answer Questions (complete the sentence), quiz, brief tests and questionnaires	28.9%	13.2%	13.2%	10.5%
• Essays	26.3%	13.2%	13.2%	2.6%
• Oral examinations <u>with checklist</u>	21.1%	10.5%	10.5%	5.3%
• Oral examinations <u>without checklist</u>	21.1%	5.3%	5.3%	2.6%
• Research projects	18.4%	5.3%	5.3%	2.6%
• Poster or other presentations	13.2%	15.8%	0%	0%
• Observation of students' performance of technical skills in preclinical labs	31.6%	0%	0%	0%
• Observation of students' performance in computer-based or other types of simulations	15.8%	2.6%	2.6%	0%
• Chart-stimulated oral exams (clinical reasoning examinations)	10.5%	2.6%	0%	2.6%
• Objective Structured Clinical Examinations (OSCE)	52.6%	52.6%	18.4%	15.8%
• Structured Clinical Operative Test (SCOT)	18.4%	2.6%	0%	0%
• Triple Jump	7.9%	5.3%	2.6%	2.6%
• Structured Workplace-Based Assessment of students' performance (WPBA) <u>termly</u>	15.8%	7.9%	10.5%	7.9%
• Structured Workplace-Based Assessment of students' performance (WPBA) <u>every patient</u>	31.6%	21.1%	18.4%	15.8%
• Structured Workplace-Based Assessment of students' performance (WPBA) <u>occasionally</u>	21.1%	15.8%	13.2%	15.8%
• Students self-assessment	26.3%	18.4%	15.8%	13.2%
• Students peer-assessment	10.5%	10.5%	7.9%	7.9%
• Assessment by patients	7.9%	7.9%	2.6%	2.6%
• Portfolios	21.1%	13.2%	13.2%	13.2%
• Learning diaries, log books, record review	18.4%	5.3%	10.5%	7.9%
• Quantity / Number of procedural units performed by students	15.8%	0%	0%	0%
• Other Assessment Methods				
Long case using virtual patients	2.6%	2.6%	0%	0%
Comprehensive incident report	0%	2.6%	0%	0%

Remarkably, the latter was also frequently used to assess each ones' communication skills (18.4%) and was third after OSCE (52.6%) and WPBA (21.1%) in this domain.

Regarding professionalism, the most frequent practices were WPBA with every patient and OSCE, both being used by the same number of institution (18.4%).

Management and leadership was equally assessed (15.8%) by OSCE, WPBA with every patient, and occasional WPBA. The least frequently used tools from the proposed list were the triple jump (2.6%) and assessment performed by patients (2.6%). Further, only two assessment methods not contained in the survey list were reported by one institution each: Long case using virtual patients for clinical and communication skills (2.6%), and Comprehensive incident report for communication skills (2.6%).

2.6 Discussion

Collating the different teaching and assessment approaches used at KCLDI required a close scrutiny of all the programme coursebooks which provided an excellent opportunity to understand how the curriculum was both organised and constructively aligned (Biggs and Tang, 2011).

The KCLDI BDS programme embraces a wide variety of teaching and assessment methodologies of knowledge, skills and attitudes/behaviours learning outcomes across all four GDC defined domains. On this basis, the first hypothesis of this Chapter can be accepted: KCLDI encompass the teaching and assessment of knowledge, skills and attitudes/behaviours in all four GDC domains (Clinical, Communication, Professionalism, and Management and Leadership).

Some other more specific findings from this Chapter helped to inform the studies in the later sections. An example of this is the balance between “lectures” (40%) and “practicals and clinical” (38%) teaching sessions (Figure 2-3). Lectures are intended to develop student’s knowledge and are mostly utilised at the beginning of the programme (Table 2-4) and are supported by a larger number of tutorial and workshop sessions aimed at increasing students’ understanding of the given knowledge (King's College London Dental Institute, 2012). The subsequent practical and clinical sessions, intended to expand students’ dexterity and clinical skills, are logically concentrated towards the end of the programme.

Both “King’s Learning, Teaching and Assessment Strategy” (King's College London Academic Services, 2011) and the GDC “Preparing for Practice” (General Dental Council, 2012a) guidelines, task dental teaching institutions to prepare increasingly

creative, critical, analytical, reflective and self-directed learner students. Given this, the reported 5% “self-directed learning” teaching sessions might be seen as lower than expected or perhaps desirable and it may be sensible to look to increase this element to facilitate further development of the students higher orders thinking skills (Spronken-Smith and Walker, 2010; Speyer *et al.*, 2011). Similarly, the introduction of assessment tools that promote reflection, critical thinking and continued learning like self- and/or peer-assessment and portfolios, could be considered (Manogue *et al.*, 2011).

A review of the learning outcomes that shape the KCLDI 2011/2012 BDS programme against the new GDC taxonomy (Table 2-1), shows an uneven combination of “Knowledge” (53.6%), Skills (32.2%) and Attitudes/Behaviours (14.2%) (Table 2-5). However and debatably, 58.6% of the Clinical learning outcomes represent knowledge (58.6%) followed by skills (38.2%) and to a much lesser extend Attitudes/Behaviours (3.2%). One would expect this Clinical domain to be a balanced combination of knowledge and skills. However, as mentioned above, the studied KCLDI 2011/2012 BDS academic programme was developed to comply with the 2002 GDC guidelines (General Dental Council, 2002) and does not represent the planned new curriculum introduced in 2013.

The distribution of KCLDI assessment methodologies within Miller’s model shows the robust base of knowledge at the “Knows” (38.6%) and “Knows How” (33.3%) levels which the Institute intends to provide its students with (King's College London Dental Institute, 2012). This is particularly true at the beginning of the programme (Figure 2-6), and allows them to develop and practice their clinical skills firstly in a simulation laboratory environment before progressing to the real workplace of the patients’ clinic. The new 2012 GDC requirements to introduce the teaching and assessment of “behaviours” besides knowledge, skills and attitudes (General Dental Council, 2012a), will certainly create the need to develop and introduce new assessment methodologies which will address the two apical levels of Miller’s model.

Within the 153 KCLDI assessment sessions, formative in-course assessment represented 55% of the total. This, as discussed in Chapter 1 (Assessment purposes in page 43), is in keeping with educational philosophies where assessment *for* learning should take priority over assessment *of* learning (Schuwirth and van der Vleuten, 2011b). The use of twenty-three assessment methodologies throughout the programme mean that each competence is assessed in more than one way and that each

methodology is also used to assess more than one competence which enables triangulation, in keeping with current best practice (Manogue *et al.*, 2002; Schuwirth and van der Vleuten, 2011b). The number and nature of assessment methodologies are in accordance with the reported twenty-one assessment methods utilised at the Bachelor of Oral Health program at Griffith University School of Dentistry (Nulty *et al.*, 2010) and the seventeen assessment strategies described to be used by fifty-three U.S. dental schools (Albino *et al.*, 2008), which further supports the process.

A striking feature of the assessment practices survey is how similar the assessment strategies are of dental teaching institutions from different countries and continents. This might be interpreted as a clear sign of the growing international adoption and convergence of higher education practices and models at work today (Collini, 2012 pages 14-15). Additionally, the fact that all the assessment methods listed in the survey were used by at least one institution, together with the point that only two other methods were reported, subscribe to the content validity of the survey itself (Streiner and Norman, 2008).

The most common assessment method reported by the 38 surveyed institutions was the OSCE (52.6%). In contrast, OSCE only represented a 3% among other 17 assessment method used by 53 surveyed US dental schools (Albino *et al.*, 2008). This despite the fact that this multi-competency examination was introduced in Scotland by Harden *et al.* (1975) almost 40 years ago. Possible explanations for the low usage of the OSCE in American dental schools might be the fact that the US survey was undertaken over 6 years ago or perhaps that the assessment of “Shows how” third level of Miller’s pyramid framework (Miller, 1990), is undertaken through other methods e.g. laboratory practical’s (8%), self-assessment (7%), unit requirements (3%), computer-based simulations (3%) and chart-stimulated evaluation (2%).

When combining the traditional multiple choice questions linked and not-linked to cases in the clinical domain, both our results and those of Albino *et al.* (2008) show these to be the most commonly used assessment technique (61% and 45%, respectively). This provides further evidence of the current international convergence of higher education practices and models in relation to assessment of the “knows” (factual recall) and “knows how” (apply biomedical information) levels of competence in Miller’s pyramid (Miller, 1990; Wass *et al.*, 2001; Shumway and Harden, 2003).

The fact that in the current study only 10.5% of the surveyed dental teaching institutions

reported the use of peer-assessment for the clinical and communication domains and a lower 7.9% for the professionalism, and Management and leadership domains, denotes that the second hypothesis of the study must be rejected: Peer-assessment is employed in more than 50% of the surveyed (inter)national dental teaching institutions.

However, this proportionally low usage of peer-assessment should not be taken negatively as it is eminently achievable and provides an opportunity to foster our students' achievements (Reported benefits of peer-assessment in page 54).

Given the role and responsibility of dental education providers to prepare students to be creative, critical and analytical, and to carry out reflective practice and self-directed learning so that they become life-long learners (King's College London Academic Services, 2011; General Dental Council, 2012a), the number of assessment methods that adhere to the American Dental Education Association (ADEA) suggestions to assess critical thinking and problem solving are small. Thus, according to Kramer *et al.* (2009) the preferred assessment method for these competencies is the "structured essays", whilst "structured observation" is felt to be only acceptable. Our findings support the idea that only a few assessment methods are related to students' higher order thinking skills as in the qualitative phase of the SOLO taxonomy (Biggs and Tang, 2011 pp. 90-91). In our experience, these are essays and clinical reasoning examinations (Chapter 4), and peer-assessment (Chapter 5).

On this subject of encouraging students' reflection, it is important to consider again the work of Biggs and Tang in that "self- and peer-assessment are particularly helpful for training students to reflect on the quality of their own work" (Biggs and Tang, 2011 p. 196). These indications and the positive effects they have on students' learning have also been highlighted by other education scholars such as Dochy *et al.* (1999) and Boud and Falchikov (2007). Thus, the wide spread use of self-assessment in all four domains reported in the present study survey is encouraging. Further, although to a lesser degree, some institutions reported the use of peer-assessment (Table 2-6). In contrast, US dental schools in one study did not report the use of peer-assessment and they only used self-assessment (Albino *et al.*, 2008). However, a second study one year later (Haden *et al.*, 2010) on "Curriculum Change" which surveyed 50 US and 5 Canadian dental schools, reported the "clinical student-faculty group practice teams" to mimic the real working environment, as the most frequently mentioned completed (64%) and in progress (19%) innovation in how pre-clinical and clinical dental education was currently being

delivered. Putting both these American studies together (Albino *et al.*, 2008; Haden *et al.*, 2010), and given the potential of a combined peer-assessment followed by self-assessment / reflection strategy (Light *et al.*, 2009 pp. 221-223) (Chapter 1, page 54), it could be interpreted as a missing opportunity to foster students' academic and reflective skills.

2.7 Conclusion

This Chapter details the variety of teaching and assessment methodologies used at KCLDI and in a wider dental community of 38 institutions from around the world. It also provides insight and describes how a five-year dental curriculum is not only theoretically organised but also delivered.

The findings demonstrate that in addition to traditional assessment methods, a number of newly introduced clinical assessment techniques, more frequently described in the medical literature (Swanwick, 2011), are also being employed in undergraduate dental education. Some of these have good education evidence of their benefit to future learning (Topping, 1998; Boud and Falchikov, 2007 pp. 114-143; Biggs and Tang, 2011 pp. 217, 245-246, 266-267). One example of this is peer-assessment although some studies limit its use to communication and interpersonal skills and health promotion competencies, highlighting especially its limits by the peers' level of knowledge and ability (Arnold, 2002; Sluijsmans *et al.*, 2002; Nilson, 2003; Norcini, 2003a). However, the evaluation detailed in the following chapters challenges this position and demonstrates that undergraduate dental students can effectively peer-assess each other in all four GDC domains: Clinical, Communication, Professionalism, and Management and Leadership.

Chapter 3

Developing and Piloting a Structured Protocol for Dental Students' Prospective Peer-Assessment and Peer-Feedback

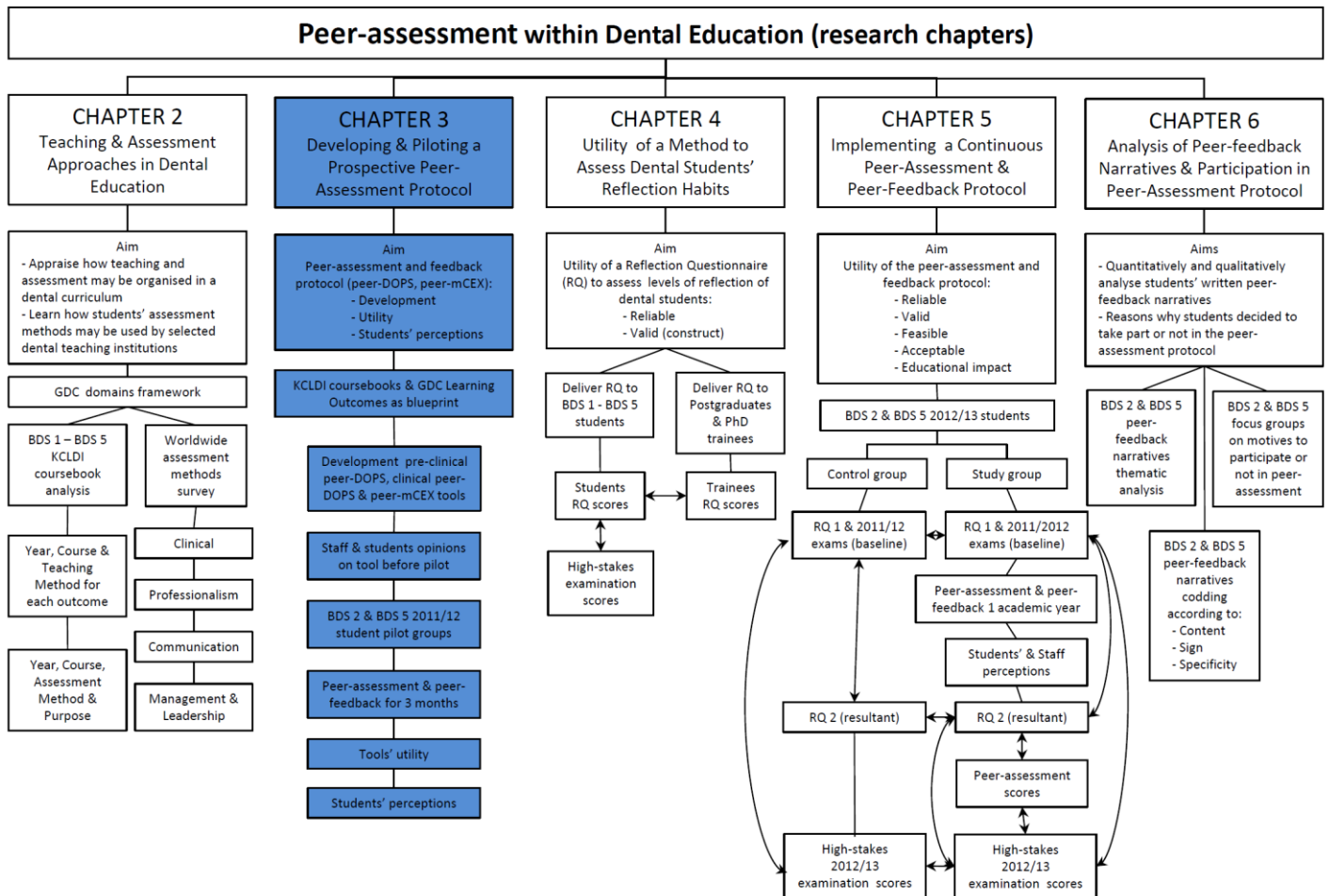


Figure 3-1 Flowchart of the study's research chapters. The third chapter is highlighted to help in providing an overall view.

3.1 Introduction

As described in Chapter 1, during the last decade or so, the use of peer-assessment in dental education has been encouraged (page 52). Consequently, recent studies on dental students have reported different peer-assessment methods, both in pre-clinical and clinical settings, with the purpose of evaluating mainly its reliability.

From the studies discussed in Chapter 1, it is known that peer-assessment scores, as compared to experienced staff marks, are reliable when pre-clinical students assess each

others' technical skills (Satterthwaite and Grey, 2008; Karl *et al.*, 2011; Taylor *et al.*, 2013), and simulated clinical interprofessional communication and examination skills (Ogden *et al.*, 2000). In contrast to the latter study, Lanning *et al.* (2011) found no correlation between peer-and staff scores in communication skills.

Further, and in agreement with the education literature (Topping, 1998), it has also been reported that dental students self-assessment and peer-assessment marks differ widely (Larsen and Jeppe-Jensen, 2008), and that peer scores reflect trainer ones more accurately (Evans *et al.*, 2007).

Notwithstanding the contribution of these studies, there is still much to know about dental students peer-assessment. It is arguable that despite providing detailed information on the marking phase of the peer-assessment procedures, none of the above reports offered any information on the “learning” usefulness of the experience. Furthermore, it is not known whether dental peers can reliably assess one another in other domains of their practice such as knowledge, skills, professionalism, and management.

In consideration of the above and given the previous chapter's reported low usage of peer-assessment by surveyed dental teaching institutions, it was deemed necessary to primarily develop and pilot (current Chapter), subsequently implement and psychometrically analyse (Chapters 5 and 6) a peer-assessment protocol especially designed for undergraduate dental students peer-assessment at the top two levels of Miller's pyramid (Miller, 1990), that could take advantage of a three-pronged strategy. That is, the enhanced potential of a combined prospective peer-assessment and peer-feedback, followed by a self-assessment and reflection phases (Schön, 1983; Ashley *et al.*, 2006; Light *et al.*, 2009 pp. 221-223). All this should contribute to the students' formative phase of learning (assessment *for* learning) and help to develop a rounded self-directed, life-long learner and reflective dental practitioner (American Commission on Dental Accreditation CODA, 2010; Australian Dental Council, 2010; General Dental Council, 2012a).

3.2 Aim

The aim of the study in this Chapter was to develop and pilot a structured protocol of formative, prospective peer-assessment of undergraduate pre-clinical and clinical dental students' skills in all four clinical, communication, professionalism, and management

and leadership GDC domains, which would be used as an informed framework for the provision of immediate peer-feedback.

3.3 Hypotheses:

- i. The developed prospective peer-assessment protocol reliably allows pre-clinical dental students to identify competence differences in various domains.
- ii. The developed prospective peer-assessment protocol reliably allows clinical dental students to identify performance differences in various domains.
- iii. Most students have positive perceptions of the peer-assessment and peer-feedback protocol.

3.4 Materials and Methods

3.4.1 Developing the instruments

Two widely used workplace-based assessment forms designed for the trainer assessment of medical/dental students and trainees behavioural levels at the top of Miller's pyramid (1990) in a naturalistic setting (Boursicot *et al.*, 2011), namely Direct Observation of Procedural Skills (DOPS) (Appendix 6) and the Mini-Clinical Evaluation eXercise (Mini-CEX) (Appendix 5) (Norcini and Burch, 2007; The Royal College of Surgeons of England, 2014), were used as starting point templates to develop three new peer-assessment instruments.

The reason behind choosing DOPS and Mini-CEX as inspiring frameworks was based on the aims of the study of developing peer-assessment tools which allowed students to peer-assess and provide face-to-face immediate dialogic peer-feedback in all clinical, communication, professionalism, and management and leadership GDC domains. The mini-Peer Assessment Tool (mini-PAT) (Appendix 10) was discarded as it is intended for anonymous peer-assessment of students' overall humanistic performance (Archer *et al.*, 2008; Norcini, 2011). The newly developed forms were designed as a framework for a continuous and structured peer-assessment and peer-feedback protocol of undergraduate dental students' specific procedures pre-clinical competence and clinical performance. A pre-clinical peer-DOPS for peer-assessment of any training procedure performed at the simulation skills laboratory, along with a clinical peer-DOPS for the

peer-assessment of whichever clinical procedure students performed on their patients, and a clinical peer-mCEX for the peer-assessment of first time patients' clinical examinations/assessment (new patients coming to the clinic), were developed.


Both pre-clinical and clinical peer-DOPS and the clinical peer-mCEX forms included a new general layout, which initially kept the traditional norm-referenced assessment scale of Below expectation, Borderline, Meets expectations and Above expectations (Norcini, 2011) as shown in Figure 3-2.

Dental Institute

Peer-Assessment & Peer-Feedback

Direct Observations of Procedural Skills (DOPS)

(Please complete this Form using a cross (X) and CAPITAL LETTERS)



University of London

I. Assessor student's name: _____

II. Assessor's KCL N° _____ III. Date: _____

IV. Have you received training as assessor in Workplace Based Assessment: ☐ Yes ☐ No

V. How many times have you assessed previous DOPS with any trainee: ☐ 0 ☐ 1 ☐ 2-5 ☐ 6-9 ☐ >9

VI. Trainee student's name: _____

VII. Trainee's KCL N° _____ Trainee's role: ☐ Assistant ☐ Dentist

VIII. Number of times this procedure has been performed by trainee: ☐ 0 ☐ 1 ☐ 2-5 ☐ 6-9 ☐ >9

IX. Case Complexity: ☐ Low ☐ Moderate ☐ High X. Time assessor observing: _____ minutes.

XI. Please grade the following areas using the given scale:	Below expectation		Borderline	Meets expectations		Above expectations	
	1	2	3	4	5	6	

Figure 3-2 Initial header of the peer-assessment DOPS form which maintained the traditional norm-referenced assessment scale.

However, while preparing the training session to be delivered to students (explained below), it was soon noticed this scale, designed for trainer assessment of trainees, would have been difficult for undergraduate students to use as it was understood they might not have the experience/knowledge to judge where their peers' work was in relation to the "norm". In the same way, Norcini (2003a) had already reported this norm-referenced scale might prove difficult for junior doctors to make judgements about their peers' performance quality. Consequently, taking the example of The Royal College of Surgeons of England (The Royal College of Surgeons of England, 2014), a criterion-referenced scale containing four written descriptions based on the needed frequency of clarification, replaced the norm-referenced scale (Figure 3-3). An "unable to comment" option when a given behaviour was not observed was also included.

Dental Institute
Peer-Assessment & Peer-Feedback
Direct Observations of Procedural Skills (DOPS)

(Please complete this Form using a cross (X) and CAPITAL LETTERS)



- I. Assessor student's name: _____
- II. Assessor's KCL N° _____ III. Date: _____
- IV. Have you received training as assessor in Workplace Based Assessment: ☐ Yes ☐ No
- V. How many times have you assessed previous DOPS with any trainee: ☐ 0 ☐ 1 ☐ 2-5 ☐ 6-9 ☐ >9
- VI. Trainee student's name: _____
- VII. Trainee's KCL N° _____ Trainee's role: ☐ Assistant ☐ Dentist
- VIII. Number of times this procedure has been performed by trainee: ☐ 0 ☐ 1 ☐ 2-5 ☐ 6-9 ☐ >9
- IX. Case Complexity: ☐ Low ☐ Moderate ☐ High X. Time assessor observing: _____ minutes.

XI. Please <u>confidentially</u> grade trainee's performance on the following areas using the given scale:	Required <u>frequent</u> clarification, warning and/or assistance	Required <u>some</u> clarification, warning and/or assistance	Required <u>very little</u> clarification, warning and/or assistance	Required <u>no</u> clarification, warning and/or assistance	I have not observed and feel Unable to Comment
--	---	---	--	---	--

Figure 3-3 Header of the peer-assessment DOPS form containing now the criterion-referenced assessment scale.

In order to contribute to the content validity of each of the three new instruments, the most relevant learning outcomes from the respective pre-clinical and clinical courses were selected to represent the corresponding assessment scales (Streiner and Norman, 2008). Based on blueprinting principles (Crossley *et al.*, 2002b), pre-clinical peer-DOPS scale represented the main learning outcomes of the BDS 2 coursebooks (Table 3-1). Similarly, clinical peer-DOPS (Table 3-2) and peer-mCEX (Table 3-3) forms represented both BDS 5 coursebooks learning outcomes, as well as those learning outcomes students must be able to demonstrate by the end of their training in order to register with the GDC as a dental professional (General Dental Council, 2012a).

After several drafts, five internal pre-clinical and clinical teachers (each of whom had at least seven years of teaching experience) ensured both pre-clinical and clinical peer-DOPS as well as clinical peer-mCEX forms sampled all the relevant and pertinent domains and appeared appropriate for the intended purpose (Streiner and Norman, 2008).

Table 3-1 Pre-clinical peer-DOPS checking of the content validity. Each item to be assessed was present in one or more BDS 2 coursebooks learning outcomes. The GDC domains have been coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93).

Item	Peer-DOPS assessment items for BDS 2	Learning Outcomes KCLDI coursebook	GDC domain
1	Knowledge of technique and dental materials used for this procedure	- Conservative Dentistry Pre-Clinical and Clinical	Clinical
2	Preparing for procedure according to taught protocol	- Integrated Clinical Care - Conservative Dentistry Pre-Clinical and Clinical	Clinical
3	Technical skills, manual dexterity and instruments handling	- Conservative Dentistry Pre-Clinical and Clinical	Clinical
4	Working position and indirect vision	- Conservative Dentistry Pre-Clinical and Clinical - Periodontology	Clinical
5	Following sequence and completing accurately all steps of the procedure	- Integrated Clinical Care - Conservative Dentistry Pre-Clinical and Clinical	Clinical
6	Observing aseptic technique/ Infection control and safe use of instruments	- Dental Institute Infection Control - Integrated Clinical Care - Conservative Dentistry Pre-Clinical and Clinical - Periodontology	Clinical
7	Seeking help where appropriate	- Integrated Clinical Care - Conservative Dentistry Pre-Clinical and Clinical	Clinical Professionalism
8	Managing time/punctuality effectively	- Integrated Clinical Care - Conservative Dentistry Pre-Clinical and Clinical	Management & leadership
9	Supporting and communicating effectively with colleagues and tutors	- Integrated Clinical Care - Communication in Dentistry	Communication
10	Overall ability to perform procedure	- Integrated Clinical Care - Communication in Dentistry	Clinical
11	Does the trainee show insight into his/her performance?	- Conservative Dentistry Pre-Clinical and Clinical - Periodontology	Professionalism

Eventually, pre-clinical peer-DOPS forms contained 11 non-compounded items (Mackillop *et al.*, 2011b), designed for the purpose of peer-assessment of any training procedure performed at the simulation skills laboratory, while clinical peer-DOPS ones contained 13 non-compounded items. Similar to the latter one, peer-mCEX forms contained 9 non-compounded items, and were both intended for peer-assessment of whichever clinical assessment or procedure students performed on their patients. Both clinical peer-DOPS and peer-mCEX forms' items resembled very much the original forms (Norcini and Burch, 2007) and could be used by any clinical BDS student independent of her/his level of training.

Table 3-2. Clinical peer-DOPS checking of the content validity. Each item to be peer-assessed was mapped to the BDS 5 coursebooks and the "Preparing for Practice" GDC learning outcomes (2012a). The GDC domains have been coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93).

Item	Peer-DOPS assessment items for BDS 5	Learning Outcomes		GDC domain
		KCLDI coursebook	GDC 2012	
1	Demonstrates understanding of indications, dental materials, complications and technique of the procedure	- Primary Dental Care - Paediatric Dentistry - Portsmouth integrated team care	- 1.1 - 1.14 - 1.1.10	Clinical
2	Obtains informed consent after explaining procedure & possible complications	- Primary Dental Care - Oral Surgery - Paediatric Dentistry	- 1.5.3 - 1.5.4 - 3.3 - 3.4	Clinical Communication
3	Demonstrates appropriate preparation pre-procedure	- Primary Dental Care - Oral Surgery - Portsmouth integrated team care	- 1.8.1 - 1.8.5	Clinical
4	Administers effective analgesia or safe sedation	- Oral Surgery - Sedation - Therapeutics - Acute Dental Care	- 1.7.5 - 1.7.7	Clinical
5	Demonstrates appropriate technical ability in line with usual practice	- Primary Dental Care - Oral Surgery - Paediatric Dentistry - Portsmouth integrated team care	- 1.9 - 1.11- 1.14 - Overarching outcomes	Clinical
6	Demonstrates aseptic technique/Infection control & safe use of instruments & sharps	- Decontamination Skills Course - Dental Institute Infection Control	- 1.8.2	Clinical
7	Deals with unexpected events or seeks help when appropriate	- Oral Surgery - Paediatric Dentistry - Oral Surgery - Orthodontics	- 1.8.4 - 9.6	Clinical Professionalism
8	Completes post procedure managements	- Primary Dental Care - Paediatric Dentistry - Oral Surgery	- 1.7.11 - 1.8.7	Clinical
9	Communication skills (patient & team)	- Communication in Dentistry - Paediatric Dentistry - Portsmouth integrated team care - Oral Disease - Special Care Dentistry	- 3.1 - 3.2 - 4 - 5	Communication
10	Organisation/efficiency and time management	- Primary Dental Care - Oral Surgery - Portsmouth integrated team care	- 10.2 - Overarching outcomes	Management & leadership
11	Consideration of patient/professionalism	- Oral Surgery - Acute Dental Care - Dental Public Health - Portsmouth integrated team care - Special Care Dentistry	- 6 - Overarching outcomes	Professionalism
12	Overall ability to perform procedure	- Primary Dental Care - Oral Surgery - Acute Dental Care - Portsmouth integrated team care	- Overarching outcomes	Clinical
13	Does the trainee show insight into his/her performance?	- Oral Surgery - Dental Public Health - Portsmouth integrated team care	- 9.3 - Overarching outcomes	Professionalism

Table 3-3. Peer-mCEX checking of the content validity. Each item to be peer-assessed was mapped to the BDS 5 coursebooks and the "Preparing for Practice" GDC learning outcomes (2012a). The GDC domains have been coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93).

Item	Peer-mCEX assessment items for BDS 5	Learning Outcomes		GDC domain
		KCLDI coursebook	GDC 2012	
1	Interviewing/history taking skills	<ul style="list-style-type: none"> - Paediatric Dentistry - Orthodontics - Therapeutics - Acute Dental Care - Portsmouth integrated team care - Oral Disease 	<ul style="list-style-type: none"> - 1.2.1 - 3 	Clinical
				Communication
2	Physical examination skills	<ul style="list-style-type: none"> - Paediatric Dentistry - Orthodontics - Acute Dental Care - Portsmouth integrated team care - Oral Disease 	- 1.2.2	Clinical
3	Diagnostic skills and underlying knowledge base	<ul style="list-style-type: none"> - Orthodontics - Paediatric Dentistry - Acute Dental Care - Radiology - Oral Disease 	<ul style="list-style-type: none"> - 1.1.3 - 1.1.4 - 1.4.2 	Clinical
4	Communication and listening skills	<ul style="list-style-type: none"> - Communication in Dentistry - Paediatric Dentistry - Portsmouth integrated team care - Oral Disease - Special Care Dentistry 	<ul style="list-style-type: none"> - 3.1 - 3.2 - 4 - 5 	Communication
5	Clinical judgment and decision making	<ul style="list-style-type: none"> - Paediatric Dentistry - Orthodontics - Radiology - Oral Disease - Special Care Dentistry 	- Overarching outcomes	Clinical
6	Consideration for patient/professionalism	<ul style="list-style-type: none"> - Oral Surgery - Portsmouth integrated team care - Special Care Dentistry 	<ul style="list-style-type: none"> - 6 - Overarching outcomes 	Professionalism
7	Organisation/efficiency and time management	<ul style="list-style-type: none"> - Primary Dental Care - Portsmouth integrated team care 	<ul style="list-style-type: none"> - 10.2 - Overarching outcomes 	Management & leadership
8	Overall clinical competence	<ul style="list-style-type: none"> - Oral Surgery - Portsmouth integrated team care 	- Overarching outcomes	Clinical
9	Does the trainee show insight into his/her performance?	<ul style="list-style-type: none"> - Oral Surgery - Dental Public Health - Portsmouth integrated team care 	<ul style="list-style-type: none"> - 9.3 - Overarching outcomes 	Professionalism

Additionally, with the purpose of stimulating reflection on practice (Brown and Manogue, 2001), all three instruments incorporated an item on "Students' insight into their performance" to judge the trainees' self-assessment (Prescott-Clements *et al.*, 2011), which was only to be graded after providing feedback and agreeing Challenges/Actions. Finally, a 6-point Likert scale for students to rate the utility of giving/receiving feedback as a technique to improve their future performance, written

instructions on how to complete the forms and a wider explanation of the grading scale were also included.


Before starting the pilot, BDS Year 2 and Year 5 students were also asked to review the wording and content of all three forms and then use them once before feeding back. This process identified two areas of student concern. The first related to the new criterion-referenced scale and the need to grade the frequency of peer “clarification” while working. This was felt to negatively affect peer-collaboration as they would refrain from asking questions in order to obtain a better assessment. As this was not the intention of the exercise, the criterion-referenced scale was again changed this time presenting the students with two different six-option educationally-referenced scales (Figure 3-4 and Figure 3-5). The decision in favour of these types of scales was taken in the search for an anchor that was more easily understood by junior students, as several studies have reported they improve the acceptability, validity and reliability of the assessment method (Beard, 2011; Crossley *et al.*, 2011; Crossley and Jolly, 2012).

Dental Institute

Peer-Assessment & Peer-Feedback

Direct Observations of Procedural Skills (DOPS)

(Please complete this Form using a cross (X) and CAPITAL LETTERS)



I. Assessor student's name: _____

II. Assessor's KCL N° _____ III. Date: _____

IV. Have you received training as assessor in Workplace Based Assessment: ☐ Yes ☐ No

V. How many times have you assessed previous DOPS with any trainee: ☐ 0 ☐ 1 ☐ 2-5 ☐ 6-9 ☐ >9

VI. Trainee student's name: _____

VII. Trainee's KCL N° _____ Trainee's role: ☐ Assistant ☐ Dentist

VIII. Number of times this procedure has been performed by trainee: ☐ 0 ☐ 1 ☐ 2-5 ☐ 6-9 ☐ >9

IX. Case Complexity: ☐ Low ☐ Moderate ☐ High X. Time assessor observing: _____ minutes.

	Stages of Progress Achievement						U/C
	1	2	3	4	5	6	
XII. After closely watching how your peer works, please grade her/his performance on the following areas using the given scale:	Starting to develop practical ability, understanding of the subject and professional attitude	Initial ability and knowledge of the subject. Expected difficulty in completing the work. Expanding professional attitude	Satisfactory achievement and knowledge of the subject. Embedding professional attitude	More than just a rudimentary knowledge of the subject. Clear ability and achievement	Good ability, achievement and knowledge. Developing skills of analysis, self-criticism and professional attitude	Extremely good achievement. Shows initiative, analysis and self-criticism	I have not observed this and feel unable to comment

Figure 3-4 Header of the first option of an educationally-referenced scale with written explanations of progress presented to students.

Dental Institute
Peer-Assessment & Peer-Feedback
Direct Observations of Procedural Skills (DOPS)

(Please complete this Form using a cross (X) and CAPITAL LETTERS)



- I. Assessor student's name: _____
- II. Assessor's KCL N° _____ III. Date: _____
- IV. Have you received training as assessor in Workplace Based Assessment: ☐ Yes ☐ No
- V. How many times have you assessed previous DOPS with any trainee: ☐ 0 ☐ 1 ☐ 2-5 ☐ 6-9 ☐ >9
- VI. Trainee student's name: _____
- VII. Trainee's KCL N° _____ Trainee's role: ☐ Assistant ☐ Dentist
- VIII. Number of times this procedure has been performed by trainee: ☐ 0 ☐ 1 ☐ 2-5 ☐ 6-9 ☐ >9
- IX. Case Complexity: ☐ Low ☐ Moderate ☐ High X. Time assessor observing: _____ minutes.

XII. After closely watching how your peer works, please grade her/his performance on the following areas using the given scale:		I have not observed this and feel unable to comment					
	Stages of Progress Achievement						
	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 12.5%; border: 1px solid black; text-align: center;">1</td> <td style="width: 12.5%; border: 1px solid black; text-align: center;">2</td> <td style="width: 12.5%; border: 1px solid black; text-align: center;">3</td> <td style="width: 12.5%; border: 1px solid black; text-align: center;">4</td> <td style="width: 12.5%; border: 1px solid black; text-align: center;">5</td> <td style="width: 12.5%; border: 1px solid black; text-align: center;">6</td> <td style="width: 12.5%; border: 1px solid black; text-align: center;">U/C</td> </tr> </table>		1	2	3	4	5
1	2	3	4	5	6	U/C	

Figure 3-5 Header of the second option of an educationally-referenced scale with graphical explanations of progress presented to students.

These two late alternatives were much better accepted by students though they asked to merge the two anchor versions together. The resulted combined six-option educationally-referenced contained a similar graphical and shortened written anchor of the desired “increasing ability over time” (Figure 3-6). This growing capability anchor emphasised, for marking purposes, the end-of-year expectations of their respective course of training (Playford *et al.*, 2013), aimed at facilitating even more junior students’ understanding and use of the scale. For example, when a student first performs a practical task, they would be peer-rated as “starting to develop” the ability for that task. Subsequently, she or he would ideally progress to “initial capability”, followed by “constant acceptable”, “constant clear”, “constant good”, and finally “constant extremely good” ability. Consequently, the new anchor did not require students to make any judgements about the quality of performance or frequency of clarification. Further, as suggested by Pangaro and ten Cate (2013), the whole assessment framework was designed to require a reduced cognitive load from “observing” students, in order to improve their rating accuracy.

The second students' concern was the extensive general information they had to fill in at the beginning of each form. Further, they were uncomfortable with the use of the word "assessor" in relation to the "observing" student, as they were unhappy to appear "assessing" their peers. In view of this, the header of all forms was simplified and all references to "assessor" were replaced by "observer" (Figure 3-6).


Final piloted versions of pre-clinical peer-DOPS, clinical peer-DOPS and clinical peer-mCEX can be found in the Appendixes Book as Appendix 12, Appendix 13, and Appendix 14, respectively.

Dental Institute

Peer-Observation & Peer-Feedback

Direct Observations of Procedural Skills (DOPS)

(Please complete this Form using a cross (X) and CAPITAL LETTERS)



University of London

I. Date: _____ II. Course: _____ III. BDS year: _____

IV. Observer student's name: _____ V. Observer's KCL N° _____

VI. Trainee student's name: _____ VII. Trainee's KCL N° _____

VIII. Procedure to be performed: _____

IX. How many times this procedure has been performed by trainee? ☐ 0 ☐ 1 ☐ 2-5 ☐ 6-9 ☐ >9

X. Case Complexity: ☐ Low ☐ Moderate ☐ High XI. Time observing: _____ minutes.

<p>XII. After closely watching how your peer works, please grade her/his performance on the following areas using the given scale:</p>	Stages of Progress Achievement						
	<div style="display: flex; align-items: center; justify-content: space-around;"> <div style="text-align: center;"> <p>Beginning BDS 2-3-4-5 Training</p> </div> <div style="font-size: 2em;">➔</div> <div style="text-align: center;"> <p>Endpoint BDS 2-3-4-5 Training</p> </div> </div> <p style="margin-top: 5px;">Increasing ability over time</p>						
	<p><u>Starting</u> to develop</p>	<p>Show <u>initial</u> capability</p>	<p>Show <u>constant</u> acceptable competence</p>	<p>Show <u>constant</u> clear ability</p>	<p>Show <u>constant</u> good ability</p>	<p>Show <u>constant</u> extremel y good ability</p>	<p>Unable to comment</p>

Figure 3-6 Header of the six-point educationally referenced scale used in both pre-clinical and clinical peer-assessment instruments which asks the "observing" student to judge their peer's ability over time.

3.4.2 Piloting and data collection

In January 2012, 26 invited students (18 females and 8 males, aged 18 to 40, mean=24.3, sd=5.9) comprising two groups, consented to participate in this peer-assessment and peer-feedback pilot study. The first group consisted of 10 pre-clinical BDS 2 Conservative Dentistry students who were under a single clinical supervisor and working at neighbouring phantom heads. The second group comprised 16 clinical BDS 5 Primary Dental Care (PDC) students who worked on the same day of the week as

clinical partners (alternating dentist/assistant roles).

At the start of the study and following suggestions from the education (Sluijsmans *et al.*, 2002) and clinical assessment (Norcini, 2003a; Cook and Beckman, 2006; Hassell *et al.*, 2012) literature, each group of students received a 45-minute peer-assessment training and familiarisation session delivered by the same researcher (JT) relating to observation, peer-assessment, peer-feedback, action plan and completion of the instruments. Using written/video examples and role-playing, they learnt and practiced how to give (observing student) and receive (training student) confidential, brief, constructive, task-focused (Kluger and DeNisi, 1996), process-oriented (Harks *et al.*, 2013), and immediate dialogic feedback (Epstein, 2007; Miller and Archer, 2010; Finn and Garner, 2011), using their peer-assessment form domains as a framework (Beard *et al.*, 2012). BDS 2 students (organised in fixed pairs) working at neighbouring phantom heads and BDS 5 clinical partners (randomly allocated each session), acted as “observer” and “trainee”, respectively during the first half of the day and then switched roles during the second half of the day.

BDS 2 students performed their own procedures as normal while “observing” their peers’ pre-clinical work every 15 minutes to avoid interfering with their own work. BDS 5 students performed their usual clinical activities in pairs, as it is usually done in dental education (Ahmad *et al.*, 2012), so that the assistant student “observed” the dentist student while treating the patient together. The observed procedure was then used to score each of the respective pre-clinical or clinical peer-assessment forms selecting and ticking one of the six options of the educationally-referenced scale (Figure 3-6) for every domain. If a given behaviour was not observed they ticked the “unable to comment” option. These scores provided a grounded framework to engage in an informed dialogic feedback. Subsequently, they agreed an appropriate action plan to address any developmental needs (Topping, 1998; Miller and Archer, 2010). Finally, after signing the forms and placing them in a specially designed delivery box, students self-assessed and reflected based on the received feedback and noted their thoughts in a private reflection diary.

In order to investigate students’ perceptions of the prospective peer-assessment and peer-feedback protocol, during the final session of peer-assessment, both groups anonymously answered the questions presented in Table 3-4, using a 5-point Likert-scale.

Table 3-4 Anonymous questionnaire to study students' perceptions presented to both BDS 2 and BDS 5 groups at the end of the peer-assessment exercise.

To what extent do you agree that the peer-assessment and feedback protocol used in this study:	Strongly Disagree	Disagree	Nor Agree nor disagree	Agree	Strongly Agree
...assessed areas that correspond to your activity in the pre-clinic/clinic?					
...could be introduced in the future to all students at the Dental Institute as part of their pre-clinical/clinical education?					
...have helped you to identify learning needs and to improve your performance?					
...was acceptable and fair?					

3.4.3 Data analysis

All peer-assessment forms data were manually digitised by the same researcher (JT) into a spread sheet. To analyse students' peer-assessment scores, each of the six levels of "increasing ability over time" of the educationally referenced scale (Figure 3-6) was assigned a numerical value from 1 to 6. Thus, the "Starting to develop" initial stage of ability was given a score 1; the "Show initial capability" a score 2 and so on until the highest "Show constant extremely good ability" which was given a score 6. Subsequently, scores were checked for normality assumptions using histogram before carrying out any parametric analysis. The reliability of the tools scores was assessed independently using Generalizability coefficient (Cronbach and Shavelson, 2004).

Descriptive statistics were used to summarise peer-assessment scores, peer-observation time, peer-feedback time, and the utility of giving/receiving feedback. The same method was used to describe students' perceptions of the studied peer-assessment and peer-feedback protocol. When comparing various measures observed for BDS 2 and BDS 5 groups, independent samples t-test was used.

In order to study whether the developed peer-assessment tools allowed students to identify competence (pre-clinical BDS 2) and performance (clinical BDS 5) differences

in various domains, a one way Analysis of Variance (ANOVA) was carried out to compare the mean scores of each form domains, separately. Where the ANOVA showed significant results, a post-hoc analysis was carried out using Tukey's test. The total peer-assessment scores observed at various time points during the study period were compared using repeated measures ANOVA for BDS 2 and BDS 5 groups, separately.

To evaluate the criterion validity of the peer-assessment instruments, that is, their correlation with another measure of the same outcome considered as the gold standard (Streiner and Norman, 2008) (further described in Chapter 1, section Assessment principles on page 45), a Pearson correlation analysis between the mean BDS 2 and BDS 5 students' peer-assessment scores were given along the study period and their respective official high stakes end-of-year mean examination mark, was performed.

Further, to investigate a possible educational effect of the prospective peer-assessment exercise on participating students' academic performance, independent samples t-test was used to compare the high stakes end-of-year mean examination marks of the 26 BDS 2 and BDS 5 students who used the peer-assessment protocol with their 142 and 146 classmates, respectively, who did not take part in the peer-assessment study.

All statistical analyses were carried out using SPSS[®] version 21 except for the Generalizability coefficient which was calculated using the software EduG 6.1e (Neuchatel, Switzerland).

3.5 Results

In line with current best practice (Topping, 1998; Norcini, 2003a), one researcher (JT) carefully organised, delivered and monitored the whole piloting peer-assessment process. Thus, starting on February 2012 and during six fortnightly occasions for BDS 2 and seven for BDS 5, students observed, assessed and provided feedback to one another. BDS 2 students completed 57 pre-clinical peer-DOPS forms (mean=5.7 per student) whilst BDS 5 students carried out 104 clinical peer-DOPS encounters (mean=6.5 per student). Only two peer-mCEX forms were submitted in the delivery box, probably due to the advanced period of the academic year. It should be born in mind that peer-mCEX forms were designed to be used for the clinical examinations and assessment of new patients coming to the clinic. At the time this study started, students were already ending their clinical cases before their examinations, and new patients

were not regularly seen. Therefore, considering the impracticality of performing any analysis with these reduced numbers of peer-mCEX forms, it was decided not to include them in the pilot study. Nonetheless, as presented in Chapter 5 (page 154), peer-mCEX forms were effectively included in the implementation phase of this study which covered the whole academic year.

BDS 2 students were assessed with pre-clinical peer-DOPS across seven different pre-clinical procedures, ranging from composite, amalgam and temporary restorations to root canal treatments and direct veneers. BDS 5 students were assessed with clinical peer-DOPS in nineteen clinical procedures, including oral health instruction, impression, bite and face-bow registration, composite, amalgam and temporary restorations, crown, bridge and veneer preparation and cementation, root canal treatments, wax try-in and root surface debridement.

Peer-assessment scores from both groups ranged from 2 (Show initial capability) to 6 (show constant extremely good ability) (mean=5.0, sd=0.7, mode=5), and were normally distributed. Generalizability coefficients for pre-clinical BDS 2 peer-DOPS was performed as a crossed three-facet (10 trainee (t) \times 6 occasions (o) \times 11 items (i)) random-effects (students worked in fixed pairs throughout the study), while for clinical BDS 5 peer-DOPS a nested three-facet (16 trainee \times 7 occasions \times 13 items) random-effects (students worked in random pairs), was used. Thus, the Generalizability coefficient for BDS 2 was 0.62 for six encounters whereas for BDS 5 it was 0.67 for seven encounters. The variance analysis for both pre-clinical BDS 2 peer-DOPS and clinical BDS 5 peer-DOPS tools is shown in Table 3-5.

Table 3-5 Variance components estimates for each peer-assessment tools.

Variance components	BDS-2 peer-DOPS Proportion of overall variance	BDS-5 peer-DOPS Proportion of overall variance
Trainee (t)	20.7%	3.6%
Occasion (o)	8.7%	5.0%
Item (i)	4.1%	3.1%
to	6.2%	6.5%
ti	16.5%	5.7%
oi	6.9%	15.8%
toi	36.9%	60.3%

The overall mean peer-assessment scores for BDS 2 and BDS 5 groups, along with their respective peer-observation time, peer-feedback time, and utility of giving/receiving feedback are presented and compared in Table 3-6.

Table 3-6 Mean and standard deviation (sd) of the peer-observation and feedback times (minutes), overall peer-assessment score (scale 1 to 6), and students' perception of the utility of giving and receiving feedback to improve future performance (scale 1 to 6), for each of the studied groups.

Variables	BDS 2	BDS 5	<i>p</i> value of difference
Observation time	153.2 (28.2)	100.2 (21.9)	<0.0001
Feedback time	6.7 (2.9)	4.8 (1.9)	<0.0001
Overall peer-assessment score	4.8 (0.8)	5.6 (0.8)	<0.0001
Utility of giving feedback	4.8 (1.0)	5.0 (0.6)	0.34
Utility of receiving feedback	5.1 (0.8)	5.3 (0.5)	0.19

The mean peer-assessment scores for each of the eleven BDS 2 peer-DOPS domains (Table 3-7) showed significant differences ($F=3.94$, $p<0.0001$) between these 11 items, with the post-hoc analysis using Tukey's test revealing the better performed "Observing aseptic technique ..." (item 6) was significantly ($p=0.04$) different from all other items. Similarly, BDS 5 mean peer-assessment scores for the thirteen peer-DOPS items (Table 3-8) also differed significantly ($F=6.55$, $p<0.0001$). Peer's scores for "Consideration of patients/ professionalism" (item 11) were statistically higher than all other items ($p=0.02$).

Table 3-7 Mean and standard deviation (sd) mark for each of the eleven pre-clinical peer-DOPS items showing the ability of pre-clinical BDS 2 participating students to identify differences ($F=3.94$, $p<0.0001$) in their peers' performance (57 completed forms).

Item	Pre-clinical peer-DOPS assessment items for BDS 2	Mean (sd)
1	Knowledge of technique and dental materials used for this procedure	4.7 (0.7)
2	Preparing for procedure according to taught protocol	4.7 (0.6)
3	Technical skills, manual dexterity and instruments handling	4.8 (0.7)
4	Working position and indirect vision	4.8 (0.7)
5	Following sequence and completing accurately all steps of the procedure	4.8 (0.6)
6	Observing aseptic technique/ Infection control and safe use of instruments	5.3 (0.7)
7	Seeking help where appropriate	4.8 (0.9)
8	Managing time/punctuality effectively	4.7 (0.8)
9	Supporting and communicating effectively with colleagues and tutors	5.0 (0.8)
10	Overall ability to perform procedure	4.8 (0.8)
11	Does the trainee show insight into his/her performance?	4.5 (0.8)

The prospective peer-DOPS marks of every fortnightly assessment occasion for pre-clinical (BDS 2) and clinical (BDS 5) students are shown in Figure 3-7. The repeated measures ANOVA of total peer-DOPS scores observed at the various time points (occasions), showed that the overall peer-assessment scores differed significantly between occasions ($p<0.0001$) for both BDS 2 and BDS 5 groups.

Table 3-8 Mean and standard deviation (sd) mark for each of the thirteen clinical peer-DOPS items showing the ability of clinical BDS 5 participating students to identify differences (F=6.55, $p<0.0001$) in their peers' performance (104 completed forms).

Item	Clinical peer-DOPS assessment items for BDS 5	Mean (sd)
1	Demonstrates understanding of indications, dental materials, complications and technique of the procedure	5.4 (1.0)
2	Obtains informed consent after explaining procedure & possible complications	5.6 (0.6)
3	Demonstrates appropriate preparation pre-procedure	5.6 (0.7)
4	Administers effective analgesia or safe sedation	5.7 (0.6)
5	Demonstrates appropriate technical ability in line with usual practice	5.5 (0.6)
6	Demonstrates aseptic technique/Infection control & safe use of instruments & sharps	5.4 (1.3)
7	Deals with unexpected events or seeks help when appropriate	5.4 (1.0)
8	Completes post procedure managements	5.3 (0.8)
9	Communication skills (patient & team)	5.4 (0.8)
10	Organisation/efficiency and time management	5.7 (0.6)
11	Consideration of patient/professionalism	6.0 (0.4)
12	Overall ability to perform procedure	5.7 (0.6)
13	Does the trainee show insight into his/her performance?	5.7 (0.5)

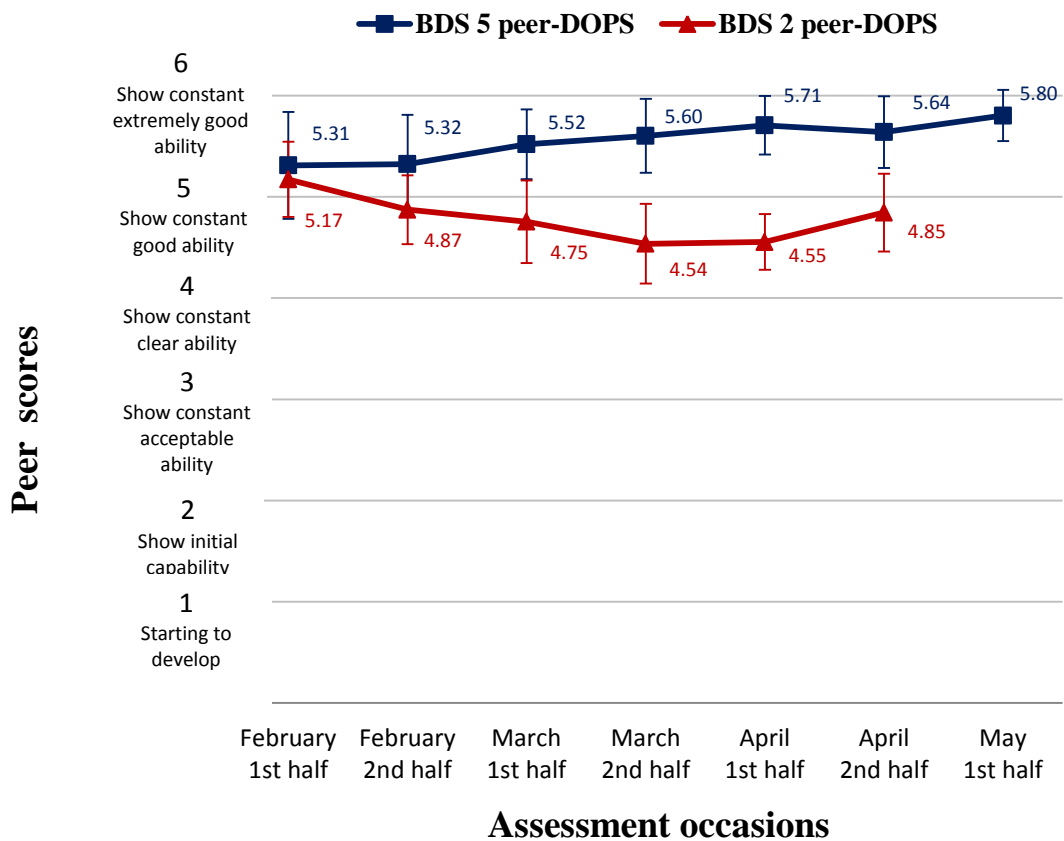


Figure 3-7 Graphical representation of prospective peer-DOPS marks for every fortnightly assessment occasions (mean and standard deviation).

Regarding criterion validity, mean peer-DOPS scores of the 10 individual BDS 2 participating students showed a positive correlation ($r=0.593$) with their high stakes end-of-year examination marks. However and despite this positive correlation, this was not statistically significant ($p=0.071$). Similarly, for the 16 BDS 5 participating students the correlation with their high stakes end-of-year examination marks was also positive ($r=0.505$) and likewise, it was not statistically significant ($p=0.051$).

Subsequently, to evaluate the possible educational impact of the prospective peer-assessment protocol, all participating students' mean high stakes end-of-year examination marks were compared to those examination marks of the rest of their respective classes. While BDS 2 high stakes examinations consisted of written short answer questions and short note questions, online multiple choice questions, and a combined clinical skills examination viva and practical, BDS 5 ones involved an essay on clinical scenarios, online short answer questions, short note questions and multiple choice questions, OSCE, clinical reasoning examination, and an oral case presentation. Thus, as shown in Table 3-9, both BDS 2 and BDS 5 studied groups did not show a statistically significant difference in their high stakes end-of-year examination marks compared to their classmates who did not use the peer-assessment protocol.

Table 3-9 Comparison of the BDS 2 and BDS 5 peer-assessment participating students high stakes end-of-year examination marks with those of their respective classmates who did not take part in the study.

Group of students	N	Mean end-of-year examination mark (sd)	<i>p</i> value of difference
BDS 2 peer-assessment group	10	60.5 (7.2)	0.886
BDS 2 All non-participating	142	60.1 (8.5)	
BDS 5 peer-assessment group	16	67.9 (5.5)	0.094
BDS 5 All non-participating	146	65.5 (5.9)	

Selected students' peer-DOPS feedback comments and the subsequent agreed action plans for both study groups are presented in Table 3-10. Further, the majority of BDS 2 and BDS 5 students expressed a positive perception of the prospective peer-assessment and peer-feedback protocol as presented in Table 3-11.

Table 3-10 Selected BDS 2 and BDS 5 students' feedback comments and agreed challenges and actions, extracted from their peer-DOPS completed forms.

	Feedback comments from peers	Agreed challenges and actions
BDS 2	Good work, maybe more packing could be used with amalgam	Better condensation of amalgam
	Shape and anatomy of amalgam must be improved	Better / improved fissure pattern
	Need specific targeted advice for improving positioning for vision in phantom head – tending to working outside head, freehand.	Work more together with demonstrator to facilitate affective practice in the phantom head
	Difficulty with indirect vision e.g. Direct for posteriors	Practice using dental mirrors to get used to indirect vision
	Margin of restoration could have been smoother	Cavity margins smoother and better anatomy on next restoration
	Time management lacking	To be more efficient, aim to two teeth per session
BDS 5	Very difficult case. Could be better listening to patient and own time management	In the future try not to undertake work beyond own capabilities and ensure complexity of dentistry is understood
	Being more confident in own ability	Correct laboratory instructions
	Trouble placing temporary crown on with Temp Bond	Better control throughout
	Problem placing rubber dam. Place the hole more centrally for full coverage	Improve technique
	Could improve infection control procedures, technically and clinically	Better cross infection control
	Student a little out of practice in jaw registration of partial dentures	Read up on jaw registration of partials

Table 3-11 Anonymous BDS 2 (n=10) and BDS 5 (n=16) students' perceptions (%) of the peer-assessment protocol after using the peer-DOPS instruments for 6 and 7 occasion, respectively.

To what extent do you agree that the peer-assessment and feedback protocol used in this study:	Course	Strongly Disagree	Disagree	Nor Agree nor disagree	Agree	Strongly Agree
...assessed you in areas that correspond to your activity in the pre-clinic/clinic?	BDS 2	0	0	0	70.0	30.0
	BDS 5	0	0	0	37.5	62.5
...could be introduced in the future to all students at the Dental Institute as part of their pre-clinical/clinical education?	BDS 2	0	0	20.0	40.0	40.0
	BDS 5	0	0	12.5	12.5	75
...have helped you to identify learning needs and to improve your performance?	BDS 2	0	0	30.0	50.0	20.0
	BDS 5	0	0	12.5	50.0	37.5
...was acceptable and fair?	BDS 2	0	0	20.0	50.0	30.0
	BDS 5	0	0	0	18.8	81.3

3.6 Discussion

The present Chapter reports the development and piloting of a structured protocol of formative prospective peer-assessment of undergraduate pre-clinical and clinical dental students' skills, used as an informed framework for the provision of immediate peer-feedback using two newly generated peer-assessment tools.

Accordingly, BDS 2 and BDS 5 students used direct observation as a method of collecting information, guided by their peer-assessment tools as micro-frameworks (Pangaro and ten Cate, 2013), to make judgements of their peers' process of simulation exercises (pre-clinical students) and patient care (clinical students). In contrast to what staff trainers do, they did not judge their peer's competence or performance "quality" nor the practice "volume" (Norcini, 2003a).

Both BDS 2 and BDS 5 overall mean peer-DOPS scores (4.8 and 5.6, respectively, Table 3-6) were higher than the reported 4.55 (sd=0.48) by Archer *et al.* (2008) when 553 medical Foundation trainees were peer-assessed once using the Mini-PAT (25% were assessed twice). Despite these differences in means, peer-scores ranged from similar minimum values of 2 and 2.13, to maximum ones of 6 and 5.77, for our and their study, respectively. However, this was not the case in the study by Bennett *et al.* (2012), where 40 undergraduate medical students assessed each other's diagnostics and/or treatment plan skills using a standard Mini-CEX form and reported no students' scores lower than 4 out of 5, which they stated was unlikely to accurately reflect the range of their students' performance.

The reliability coefficient G was higher for BDS 5 clinical peer-DOPS (0.67 for seven occasions) as compared to BDS 2 pre-clinical peer-DOPS (0.62 for six occasions). These G coefficients are comparable to those reported by Wilkinson *et al.* (2008) when six Multi-Source Feedback (MSF) encounters (0.65) were used to assess medical specialists. However, they are lower than those reported by Moonen-van Loon *et al.* (2013) for six (0.74) and seven (0.76) MSF residents' encounters performed by supervisors, peers, nurses, administrative staff, patients, and self-assessment. Further, they are also lower than the generally accepted G coefficient of ≥ 0.7 as sufficient for 'low-stakes' assessment situations (Beard *et al.*, 2011). A possible explanation for this might be found in the number of peer-assessment encounters completed by the students. BDS 2 only reached a mean of 5.7 encounters per students, while BDS 5 reached a mean of 6.5 per students. Both these cases are lower than the minimum of 7 to 11

judgements required for reliable findings as concluded by Williams *et al.* (2003). A longer and larger peer-assessment study using the current protocol might be needed to evaluate the effect of more students' encounters on the reliability of the instruments.

Despite both groups of undergraduate students being on the path from novice to experts (Dreyfus *et al.*, 1988), and against the Sluijsmans *et al.* (2002) hypothesis that novices are less capable of assessing certain domains, the analysis of the current data showed that both groups were able to identify differences in competence and performance when assessing their peers across the respective eleven and thirteen domain peer-DOPS forms. This being the case, the first hypothesis of this Chapter can be accepted: The developed prospective peer-assessment protocol reliably allows pre-clinical dental students to identify competence differences in various domains. Similarly, the second hypothesis can also be accepted: The developed prospective peer-assessment protocol reliably allows clinical dental students to identify performance differences in various domains.

Accordingly, BDS 2 students distinguished between a better clinical skill in "Observing aseptic technique..." and a lower professional behaviour in "Does the trainee show insight into his/her performance?" (Table 3-7). Whilst the former (mean 5.3) corresponds well with the intensive training students have in infection control and safety starting as early as BDS 1, the latter (mean 4.5) was not expected. Thus, despite the natural motivation for "friendship" marking as described by Norcini (2003b), peers scored each other significantly low ($p=0.04$) in this professionalism item. In contrast to a possible blandness, this might be better explained by a competitive spirit as illustrated by Cushing *et al.* (2011), or as a real sign that these junior students are still developing this self-assessment skills. This BDS 2 item low scoring was not observed in the same item of senior BDS 5 students' peer-assessment. By way of contrast, BDS 5 students assessed the professional behaviour of "Consideration of patient/ professionalism" item with the maximum possible score (mean 6), demonstrating a high social commitment. At the same time, several other items in the clinical (items 1 & 6), management (item 8) communication (item 9), domains were significantly lower. These findings are in agreement with the study of Bennett *et al.* (2012), who reported medical undergraduates' peer-assessment ability to identify "areas in which peers performed well and those that required improvement".

Yet another indication of the students' ability to make an accurate inference of their

peer' performance (Streiner and Norman, 2008) was their ability to detect changes in their peers' performance with time, though with dissimilar patterns (Figure 3-7). In the case of BDS 5 students they were able to perceive significant progress in their peers' performance by the third peer-assessment session. In contrast, BDS 2 students only started to notice peer improvement at the fifth session, while during the first four encounters scores dropped significantly. This difference might be explained by an initial calibration process (Hauser and Bowen, 2009), or the time needed to gain experience as evaluators (Karl *et al.*, 2011), or even adjusting to the learning environment (Schoenrock-Adema *et al.*, 2007), before scores started to rise. This increase in ratings over time is in keeping with the earlier findings of Prescott-Clements *et al.* (2011) in postgraduate dental trainees assessed by staff members in eight similar items, and Davies *et al.* (2009) in foundation medical trainees' peer-assessment.

Further, whilst not statistically significant, both groups' peer-assessment scores were positively correlated to students' end-of-year examination marks. Once again, a larger peer-assessment study of the current protocol might be needed to better evaluate this validity criterion.

Similarly, despite students' high perceptions in that the peer-DOPS protocol "helped them to identify learning needs and improve performance" (Table 3-11), it had no observable educational impact as both groups of participating students show no difference in high stakes end-of-year examination marks when compared to the other students of their respective classes. This might be explained once again by the short observation period (around three months). Already graduated foundation medical trainees' peer-assessment (using the mini-PAT), required two consecutive semester assessments to detect a slight increase in scores from 4.6 to 4.7 (scale from 1 to 6) (Davies *et al.*, 2009).

At the end of the piloting phase of the study, an anonymous questionnaire of both BDS 2 and BDS 5 groups of students gathered their perception of the peer-assessment protocol (Table 3-11). Most of them expressed their positive perceptions of the peer-assessment and peer-feedback protocol, which allows for the third hypothesis of this Chapter to be accepted: Most students have positive perceptions of the peer-assessment and peer-feedback protocol.

Similar positive perceptions were recently reported by Kirton *et al.* (2013) when 90.8%

of 643 dental Foundation trainees, assessed by trainers, found a longitudinal evaluation of performance (D-EPs), useful.

All students (100%) agreed or strongly agreed that both peer-DOPS instruments assessed suitable domains. Likewise, the vast majority (>80%) agreed or strongly agreed that the peer-DOPS protocol was feasible to be introduced in the future to all students at the Dental Institute as part of their pre-clinical/clinical education; helped them to identify learning needs and improve performance; and probably most importantly, was acceptable and fair. The latter is particularly important as a sound assessment needs to be accepted, fair (Norman *et al.*, 1991) and authentic (McCoubrie, 2004) or it is destined to fail (van der Vleuten, 1996). The fact that students' views and opinions were requested and their answers implemented during the development phase of the peer-assessment instruments, may have contributed to an increase in their acceptability (Shue *et al.*, 2005). Further, we did not observe nor receive any negative comment or concern related to the peer-assessment experience like students not accepting peer feedback as accurate or regarding it as of little benefit, inhibited or constrained students, inappropriate and hurtful comments or any other sign of peer power relationships, as previously reported (Van Rosendaal and Jennett, 1992; Topping, 1998; Dannefer *et al.*, 2005; Boud and Falchikov, 2007).

The present pilot results are highly significant and interesting inasmuch as there is no data in the literature on the prospective use of peer-assessment in the dental undergraduate setting. However, the development and piloting of the protocol was very time-consuming and required careful and rigorous preparation as well as dedication. Further, the limiting sample size and the short observation period, means care should be exercised when trying to generalise.

3.7 Conclusion

This Chapter developing and piloting of a prospective peer-assessment protocol has established the ability of undergraduate dental students to reliably and accurately peer-assess one another in all four GDC domains. Further, it has shown that the approach was positively perceived by participating students.

However, a further implementation of the piloted peer-assessment tools over a longer period of time and across a wider study body is necessary to provide more reliable and

generalizable data. Thus, questions mainly related to the tools' utility as an assessment method (van der Vleuten, 1996), and the educational effects of the protocol's combined peer-assessment and peer-feedback, followed by self-assessment and reflection, remain unanswered. Consequently, the following chapter will describe the usefulness of a method to assess students' reflection skills in order to later, in the subsequent peer-assessment implementation phase (Chapter 5), evaluate the potential of the protocol to foster students' higher order thinking skills.

Chapter 4

Utility of a Method to Assess Dental Students' Reflection Habits

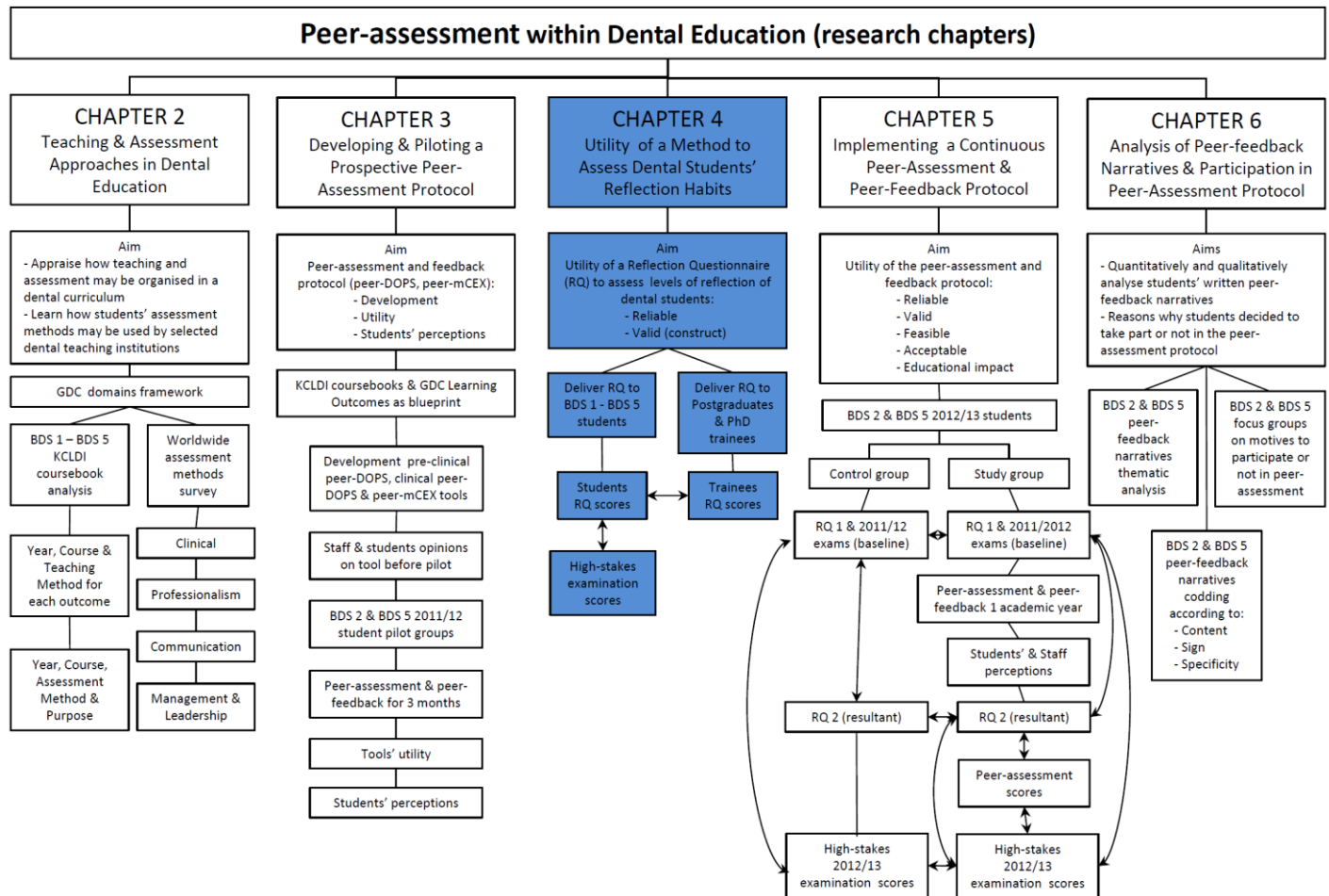


Figure 4-1 Flowchart of the study's research chapters. The fourth chapter is highlighted to help in providing an overall view.

4.1 Introduction

Reflection, as discussed in Chapter 1 (Reflection and learning in page 77), has long been acknowledged as an important concept in the learning process (Dewey, 1938; Boud *et al.*, 1985) and as on-going activity, it is regarded as a crucial and essential characteristic of healthcare professionals' education and competent professional practice (Mann *et al.*, 2009; Aronson, 2011; Chambers *et al.*, 2011).

Thus, in an educational environment, it would be useful to determine whether students

or trainees are engaged in reflective practice. This requires some means of identifying reflective thought and a measure of its depth (Kember, 1999; Pee *et al.*, 2002). However, there appears to be no current reflective practice assessment procedure with sufficient validity, acceptability or clear formulation which would qualify it as the gold standard for assessing students' reflective practice abilities (National Postsecondary Education Cooperative, 2000; Chambers *et al.*, 2011).

Consequently, several approaches have been proposed to measure students' reflectivity (Pee *et al.*, 2002). Thus, the majority of these reported methods in health sciences consist of various forms of reflective writings in medicine (Niemi, 1997; Driessen *et al.*, 2005; Learman *et al.*, 2008; Wald *et al.*, 2009; Aronson *et al.*, 2012; Dekker *et al.*, 2013); dentistry (Pee *et al.*, 2002; Boyd, 2008; Bush and Bissell, 2008; Tsang and Walsh, 2010; Jonas-Dwyer *et al.*, 2013; Tonni and Oliver, 2013a; Wallace *et al.*, 2013); nursing (Wong *et al.*, 1995); and physiotherapy (Williams *et al.*, 2002). This despite reports that students dislike "writing" reflections (Bush and Bissell, 2008), and that their academic pressure discourage them from engaging in honest and open reflection (Hargreaves, 2004),

The literature also includes accounts of self-reported questionnaires/surveys (Mitchell, 1994; Kember *et al.*, 2000; Sobral, 2000; Aukes *et al.*, 2007), reflective blogging (Wetmore *et al.*, 2010), semi-structured questionnaires with reflection-evoking case vignettes (Boenink *et al.*, 2004), surveys of attitudes and the use of reflective portfolios (Pearson and Heywood, 2004; Bush and Bissell, 2008; Kardos *et al.*, 2009; Koole *et al.*, 2013), and interviews (Hallett, 1997; Boyd, 2008).

The vast majority of these methods were designed and used as reflective learning activities in an attempt to foster students' reflective thinking, whilst the assessment and depth of the reflective thoughts followed a demanding qualitative approach. However, for the purpose of the present chapter study, a feasible method of assessing reflection skills that could be implemented in large numbers of students and trainees, both before and after the peer-assessment implementation, was required.

Further, it should also allow for comparisons of pre- and post-intervention reflection skills without influencing the sample's reflective thoughts, as well as permitting statistical correlation with other interval scales. In other words, we needed a viable, quantitative and diagnostic approach to assess reflection skills.

In view of this, we considered using the self-reported questionnaire from Sobral (2000), the Gröningen Reflection Ability Scale (GRAS) (Aukes *et al.*, 2007), and the self-reported reflection questionnaire from Kember *et al.* (2000). However, keeping in view its confirmed psychometric properties (Lethbridge *et al.*, 2011) and its extensive use in multiple disciplines (Phan, 2009; Lim, 2011; Chelliah and Arumugam, 2012; Naghdipour and Emeagwali, 2013; Ambrose and Ker, 2014), we decided that the latter was most appropriate.

4.2 Aim

The aim of this Chapter was to assess the usefulness, in terms of reliability and construct validity of the self-reported Reflection Questionnaire (RQ) (Kember *et al.*, 2000) as a method of assessing dental students' and postgraduate trainees' reflection habits.

4.3 Hypotheses

- i. The results of the studied RQ are reliable.
- ii. The studied RQ scores discriminate between subjects' non-reflective and reflective thoughts.
- iii. The levels of reflective habits from undergraduate students and postgraduate trainees are the same.
- iv. Undergraduate dental students' reflective habits and their high stakes academic achievements are positively correlated.

4.4 Materials and Methods

4.4.1 The instrument

The RQ of Kember *et al.* (2000) was developed as a simple diagnostic instrument to examine students' engagement in reflective thinking in professional academic programmes from any discipline. Thus, the questionnaire, based on Mezirow's (1991) conceptualisation of levels of reflective thinking, has 16-item (statements) that conform

four scales or factors to quantitatively assesses two levels of non-reflective actions (Habitual Action and Understanding), and two levels of reflective actions (Reflection and Critical Reflection).

According to Kember *et al.* (2008) the descriptions for each of the four categories are:

- *Habitual Action*: previously learnt responses that are automatically performed with little conscious thought. It is what expert practitioners do in routine practice as they have dealt with the same situation many times before. It is also what novice learners do when they rigidly follow the steps of taught procedures without thinking its applicability or alternatives. Habitual Action also occurs when students provide an answer without trying to understand or forming a view of the concept, task or theory that underpins the topic, which is compatible with a surface approach to learning.
- *Understanding*: a cognitive learning and/or reading but without appraising the concept or task. Still, the student attempts to reach an understanding and meaning of the theory, and therefore is related to a deep approach to learning. However, the concepts are not related to personal experiences or real applications and, as such, do not have personal meaning, may not be incorporated as new knowledge, and its retention can be limited. Understanding is usually a characteristic of inexperienced undergraduate students thinking, as they lack practice to apply the new knowledge.
- *Reflection*: an active, persistent, and careful critique of assumptions about the content or process of problem solving. It is the next level after understanding as the concepts or tasks are related to previous knowledge and personal experiences. Theory then is interpreted and practically applied.
- *Critical Reflection*: becoming aware of why we perceive, think, feel or act as we do. It implies undergoing a transformation of perspective which requires first a critical review of presuppositions and assumptions from prior learning and their consequences. These are not easy to change and therefore, critical reflection is unlikely to happen regularly, especially in professionals. However, it is expected to be more common in students who are learning a subject or profession and so are establishing their assumptions and constructing their individual meanings.

When answering the RQ, participants are asked to indicate their level of agreement with all 16-items on a five-point Likert scale ranging from “definitely disagree” (score as 1), “disagree with reservations” (score as 2), “neutral” (score as 3), “agree with

reservations” (score as 4), and “definitely agree” (score as 5) (Table 4-1).

Table 4-1 Reflection Questionnaire from Kember et al. (2000) used to assess the reflective habits of students and trainees.

Reflection Questionnaire						
	Statements	Definitely agree	Agree with reservation	Neutral	Disagree with reservation	Definitely disagree
1	When I am working on some activities, I can do them without thinking about what I am doing					
2	Dental training requires me to understand concepts taught by the lecturers					
3	I sometimes question the way others do something and try to think of a better way					
4	As a result of my dental training I have changed the way I look at myself					
5	In dental training I do things so many times that I have started doing them without thinking about it					
6	To pass dental training examinations you need to understand the content you are studying					
7	I like to think over what I have been doing and consider alternative ways of doing it					
8	Material learned during my dental training has challenged some of my firmly held ideas					
9	As long as I can remember material presented in class for examinations I do not have to think too much					
10	I need to understand the material taught by the teacher in order to perform practical tasks					
11	I often reflect on my actions to see whether I could have improved on what I did					
12	As a result of the material learned in dental training I have changed my normal way of doing things					
13	If I follow what the lecturers say, I do not have to think too much about my dental training					
14	In dental training you have to continually think about the material you are being taught					
15	I often re-appraise my experience so I can learn from it and improve for my next performance					
16	During dental training I discovered faults in what I had previously believed to be right					

Each of the four scales is measured by four contributory items, and the participant's score on each of these dimensions is calculated by adding the responses score for each of the four items from each scale. Once completed, each participant receives a score ranging from a minimum of 4 (definitely disagree in all four items of that scale) to a

maximum of 20 (definitely agree in all four items of that scale) for each Habitual Action (statements 1+5+9+13), Understanding (statements 2+6+10+14), Reflection (statements 3+7+11+15) and Critical Reflection (statements 4+8+12+16) dimensions (Table 4-1). There is no overall score as this would not be consistent with the theoretical underpinnings of the scale development (Kember *et al.*, 2000). Consequently, the higher the score in each category, the more agreement with engaging in the particular dimension that each scale measures. Lastly, the RQ, along with an Information Sheet explaining the study, a Consent Form as well as a final free text box for comments, were developed electronically and the RQ's optimal online technical performance was tested using Survey Monkey®.

4.4.2 Participants and data collection

At the end of June 2012, all 740 BDS years 1 to 5 undergraduate students and 221 postgraduate clinical dentistry programme and PhD postgraduate trainees enrolled at KCLDI, received an electronic invitation from the Director of Education or Director of Graduate Studies at KCLDI to voluntarily participate in the study by completing the online RQ.

4.4.3 Data analysis

Participation was closed twelve weeks after the invitation was sent. Completed questionnaire responses were downloaded from Survey Monkey® into a spread sheet and imported into SPSS® version 21 for analysis.

Descriptive statistics were employed to express participants' gender and age characteristics as well as their current and previous academic backgrounds. Normal distribution of the resultant RQ scores was assessed by visual inspection of a histogram and by exploring the central tendency (mean, median, and mode). The internal reliability of the items constituting each scale (Habitual Action, Understanding, Reflection and Critical Reflection) was quantified by computing Cronbach's alpha coefficient.

Construct validity of the RQ was assessed in five different ways:

- To test whether the questionnaire could distinguish between participants' four scales of reflective skills, a one-way analysis of variance (ANOVA) with post hoc analysis using Tukey's test, was performed.

- To assess the RQ capability to differentiate between participants' four scales of reflective skills according to their academic course (BDS 1 to 5, clinical postgraduate, and PhD), a one-way analysis of variance (ANOVA) with post hoc analysis using Tukey's test, was performed.
- Subjects' reflection skills for each of the RQ four scales were compared according to their gender, age group (≤ 23 years old and ≥ 24 years old) and whether they had a previous university degree, using separate independent samples t-tests.
- Inter-scale relationships between all four RQ scales, as predicted conceptually by the questionnaire developers (Kember *et al.*, 2000), were studied using a Pearson correlation analysis.
- To establish the relationship between the reflection habits of undergraduate subjects and their academic achievements a Pearson correlation analysis was performed. This was measured by collecting students' end-of-year high stakes examination marks and subsequently correlating them to their RQ scores.

To enable the fifth analysis, examination marks from Paper 1a for BDS 1 to 3 (short answer questions and short note questions), Paper 1b for BDS 4 and 5 (written essays on clinical scenarios), Paper 2 (online MCQ), an OSCE, Clinical Reasoning Examination (CRE), and Case Presentation were obtained and correlated to students' Habitual Action, Understanding, Reflection and Critical Reflection scores. It should be mentioned that all students take Papers 1 and 2 whereas the OSCE is used to assess BDS Years 3 to 5, and CRE along with Case Presentation are part of BDS Year 5 Finals.

4.5 Results

4.5.1 Participants and data collection

A total of 324 (34%) undergraduate students and post-graduate trainees (208 females and 116 males) answered the invitation and completed all 16 items of the online self-reported RQ. No issues were reported by any student or trainee and only one exited the questionnaire before completing all 16 statements. However, the same student later fully completed it in the second attempt.

Twenty-four undergraduate students submitted a free text comment at the end of the questionnaire. Two hundred and eighty-one participants were undergraduate students representing all BDS courses (years 1 to 5), while 43 were postgraduate trainees following a clinical dentistry programme (N=25) or a PhD (N=18). At the time the study was conducted, all groups of respondents had an age range between 18 and 46 years (mean=25, sd=6.7), while 44% of them were 23 years of age or younger. The number of participants who had a previous national or international university degree was 131, representing 40% of the whole sample (Table 4-2).

Table 4-2 Number of participants by course of study, mean age and standard deviation (sd) and proportion of students with a previous university degree.

Cohort	N	Mean age (sd)	Previous degree
BDS 1	38	20.5 (2.5)	10%
BDS 2	71	22.7 (4.8)	25%
BDS 3	44	24.4 (5.9)	53%
BDS 4	42	23.9 (3.0)	23%
BDS 5	86	24.9 (2.8)	39%
Clinical Postgraduate (1 st , 2 nd and 3 rd year)	25	36.5 (5.7)	100%
PhD (2 nd and 3 rd year)	18	32.4 (7.1)	100%
Totals	324	25.0 (6.7)	40%

4.5.2 Data analysis

Reflection questionnaire answers ranged from 1 (definitely disagree) to 5 (definitely agree) (mean=3.7, sd=0.9, median=4.0, mode=4) and were normally distributed. Cronbach's alpha internal reliability coefficient was 0.696 for Habitual Action, 0.711 for Understanding, 0.702 for Reflection and 0.712 for Critical Reflection.

Resultant RQ four-scale mean scores from all participants was 10.7 (sd=1.6) for Habitual Action, 17.3 (sd=1.5) for Understanding, 17.0 (sd=1.8) for Reflection, and 14.4 (sd=2.0) for Critical Reflection. The ANOVA test confirmed these scores as significantly different ($F=845.02$, $p<0.0001$). Hence, a further post hoc analysis using Tukey's test was undertaken (Table 4-3).

Interestingly, 5.8% of undergraduate students exhibited a Reflection score under two standard deviations (13.4) of the overall mean (17.0). The percentage of students lower than two standard deviations (10.2) of the overall mean for Critical Reflection (14.4) increased to 8.0%. However, no postgraduate or PhD trainees were found in these groups.

Table 4-3 Mean value for each of the four RQ scales of all subjects and ANOVA with post hoc Tukey's test results showing inter-scale mean difference and corresponding *p*-values.

Reflection Questionnaire scales									
		Habitual Action		Understanding		Reflection		Critical Reflection	
	Scale Mean	10.7		17.3		17.0		14.4	
		Mean difference	<i>p</i> value of diff.	Mean difference	<i>p</i> value of diff.	Mean difference	<i>p</i> value of diff.	Mean difference	<i>p</i> value of diff.
Habitual Action	10.7	-	-	-	-	-	-	-	-
Understanding	17.3	6.624	<0.0001	-	-	-	-	-	-
Reflection	17.0	6.322	<0.0001	0.302	0.148	-	-	-	-
Critical Reflection	14.4	3.644	<0.0001	2.980	<0.0001	2.679	<0.0001	-	-

The mean scores for each of the RQ scales according to participants' academic course and study year were calculated (Table 4-4), and graphically illustrates (Figure 4-2), grouping all undergraduate students together to compare them with clinical postgraduates and PhD trainees. The ANOVA test found no significant differences between these three groups in Habitual Action ($F=0.82$, $p=0.921$), and Understanding ($F=2.760$, $p=0.067$).

Table 4-4 Mean scores and standard deviations (*sd*) for each of the four scales of the RQ according to participants' academic course and study year.

Reflection Questionnaire scales					
Cohort		Habitual Action	Understanding	Reflection	Critical Reflection
	N	Mean (sd)	Mean (sd)	Mean (sd)	Mean (sd)
BDS 1	38	9.2 (2.1)	17.2 (1.7)	16.2 (2.0)	12.7 (2.1)
BDS 2	71	9.9 (2.3)	17.8 (1.2)	16.4 (2.1)	13.1 (2.3)
BDS 3	44	10.7 (1.7)	16.9 (1.6)	16.5 (1.7)	13.6 (1.9)
BDS 4	42	10.5 (1.5)	17.0 (1.7)	17.1 (1.9)	14.8 (1.8)
BDS 5	86	12.4 (1.4)	17.2 (2.2)	17.5 (1.3)	15.5 (1.7)
Clinical Postgraduate	25	10.3 (1.2)	17.5 (1.1)	17.6 (1.2)	16.3 (1.2)
PhD	18	10.4 (1.0)	18.5 (1.6)	18.3 (1.5)	18.1 (1.1)
Scale Mean	324	10.7 (1.6)	17.3 (1.5)	17.0 (1.8)	14.4 (2.0)

Conversely, the same test indicated significant differences between the groups in Reflection ($F=3.725$, $p=0.027$), with the post hoc Tukey's test showing undergraduate

student's (mean 17.0) and PhD trainee's (mean 18.3) scores differed significantly ($p=0.038$). However, the most significant difference was seen in the Critical Reflection scale ($F=15.312$, $p<0.0001$), where the post hoc Tukey's test revealed the differences between undergraduate student's scores (mean 14.1) and clinical postgraduate trainee's ones (mean 16.3) were significant ($p=0.036$), as well as that from undergraduates' and PhD trainees' (mean 18.1) ($p<0.0001$).

Comparisons between Habitual Action, Understanding, Reflection and Critical Reflection mean scores of the sampled subjects grouped by gender, age group and whether they had a previous university degree, found statistically significant differences (Table 4-5). Accordingly, the ≥ 24 years old group showed higher scores than those ≤ 23 years old both in Reflection (17.5 versus 16.4, $p<0.0003$) and Critical Reflection (15.1 versus 13.6, $p<0.0006$) scales. Similarly, those who attained a previous university degree exhibited higher scores those who did not in Reflection (17.5 versus 16.7, $p<0.011$) and Critical Reflection (15.3 versus 13.8, $p<0.0004$).

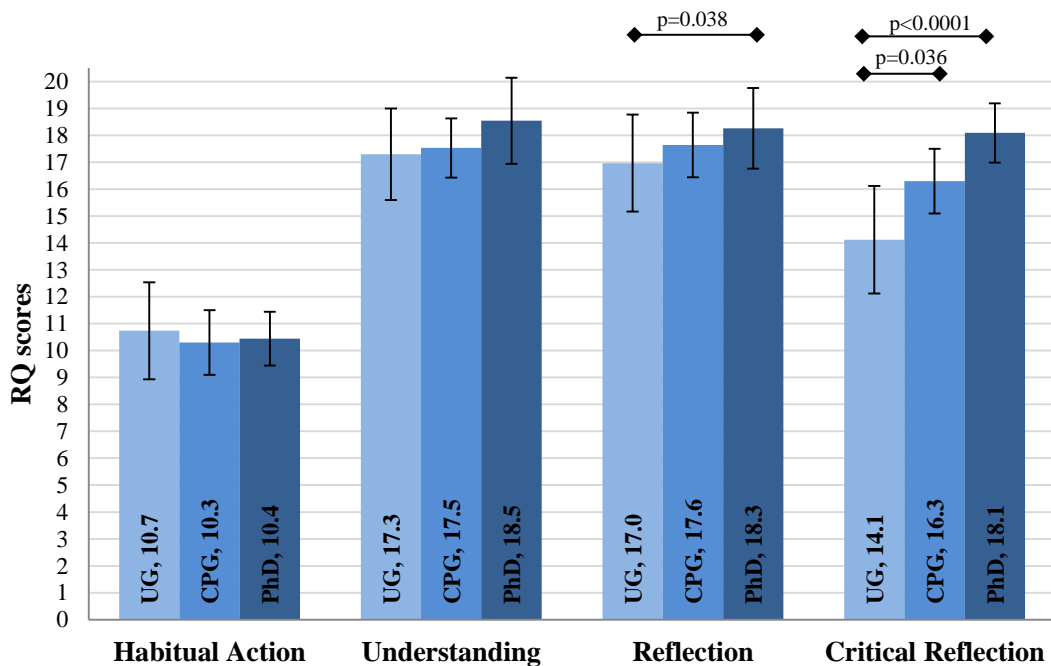


Figure 4-2 Mean score and \pm standard deviation for Habitual Action, Understanding, Reflection and Critical Reflection scales of the RQ, grouping all undergraduate students together (UG), clinical postgraduates (CPG) and PhD trainees.

Table 4-5 Separate independent samples t-test results showing number, mean scores, standard deviations (sd) and p-values for each of the four RQ scales according to participants' gender, age group and whether they had a previous university degree.

Reflection Questionnaire scales									
Cohort		Habitual Action		Understanding		Reflection		Critical Reflection	
	N	Mean (sd)	p value of diff.	Mean (sd)	p value of diff.	Mean (sd)	p value of diff.	Mean (sd)	p value of diff.
Female	208	10.7 (2.9)	0.948	17.4 (2.1)	0.476	17.0 (1.9)	0.844	14.4 (3.0)	0.786
Male	116	10.7 (2.6)		17.2 (2.0)		17.1 (2.6)		14.3 (3.2)	
≤ 23 years old	143	10.8 (2.6)	0.711	17.2 (2.0)	0.505	16.4 (2.7)	0.0003	13.6 (3.0)	0.0006
≥ 24 years old	181	10.6 (2.9)		17.4 (2.1)		17.5 (1.7)		15.1 (3.1)	
Previous degree	131	10.4 (2.7)	0.173	17.6 (1.8)	0.090	17.5 (1.7)	0.011	15.3 (2.8)	0.0004
No previous degree	193	10.9 (2.8)		17.1 (2.1)		16.7 (2.6)		13.8 (3.2)	

The tabulation of the RQ inter-scale correlations and their level of statistical significance are shown in Table 4-6 below. Understanding and Reflection showed the highest relationships, while Habitual Action and Critical Reflection exhibited the lowest.

Table 4-6 Mean value for each of the four RQ scales of all subjects and inter-scale Pearson correlations with the corresponding p-value.

Reflection Questionnaire scales									
		Habitual Action		Understanding		Reflection		Critical Reflection	
		Mean							
		10.7		17.3		17.0		14.4	
		Correlation	p value of diff.	Correlation	p value of diff.	Correlation	p value of diff.	Correlation	p value of diff.
Habitual Action	10.7	1							
Understanding	17.3	0.046	0.63	1					
Reflection	17.0	0.044	0.61	0.423	<0.0001	1			
Critical Reflection	14.4	0.035	0.69	0.400	<0.0001	0.354	0.003	1	

Correlation analysis between reflective habits of undergraduate subjects and their academic performances in all end-of-year high stakes examinations showed wide diversity of relationships between RQ scales and the different assessment programme methods and is particularly interesting in these results (Table 4-7).

Table 4-7 Correlation and corresponding p-value between each RQ scale of BDS year 1-2-3-4-5 subjects and their academic performances in end-of-year examinations. For each examination, the cohort, number of subjects and the examination mean score of the sample as well as that from the whole class () are shown.

Examination	Reflection Questionnaire scales										
				Habitual Action		Understanding		Reflection		Critical Reflection	
	Cohort	N	Exam mean scores	Correlation	P value	Correlation	P value	Correlation	P value	Correlation	P value
Paper 1a (short answer questions, short note questions)	BDS 1 to 3	153	61 (59)*	0.120	0.340	0.035	0.780	0.109	0.380	0.073	0.560
Paper 1b (written essays on clinical scenarios)	BDS 4 & 5	128	62 (61)*	0.102	0.560	0.496	0.002	0.461	0.005	0.071	0.680
Paper 2 (online multiple choice questions)	BDS 1 to 5	281	65 (66)*	0.117	0.170	0.394	<0.0001	0.145	0.120	0.049	0.520
OSCE	BDS 3 to 5	172	64 (62)*	0.446	0.016	0.291	0.003	0.138	0.250	0.184	0.120
Clinical Reasoning	BDS 5	86	65 (64)*	0.124	0.490	0.490	0.004	0.477	0.005	0.157	0.320
Case Presentation	BDS 5	86	62 (61)*	0.227	0.210	0.143	0.430	0.284	0.110	0.106	0.560

* p>0.44

Selected free text comments one from each BDS course that were submitted by participants after completing the RQ are displayed in Table 4-8, as a qualitative complement of the data presented above.

Table 4-8 Selection of participants' comments, study cohort, age, previous degree and their scores in Habitual Action (HA), Understanding (U), Reflection (R), and Critical Reflection (CR).

Comment	Study group	Age	Previous degree	HA	U	R	CR
"From the questions asked in this questionnaire, I will try and reflect more on my actions in the clinic so that I can improve the next time"	BDS 1	21	No	6	18	15	15
"Sometimes when it comes to methods, instead of thinking of better ways to do things I think to what is expected of me in exams"	BDS 2	22	No	12	19	16	11
"I keep a dentistry diary where I record what I have learned & what I can improve on in my clinical sessions and all the small bits advice I get from my tutors which are specific to the treatment situation that would not have been mentioned in the main lectures - I find this really helps build my knowledge and confidence for the next session - and also looking back on the past entries I can see how far I have already progressed as a student dentist"	BDS 3	35	Yes	5	20	19	19

“Self-assessment definitely occurs throughout the course, and reflective feedback can be very helpful, but it is in itself a learning process, and I think it definitely gets easier as you progress through the course. It is extremely helpful to have a discussion with your tutors about your performance at the end of the day, and to have a chance to explain why you did certain things a certain way - this way, misunderstandings can be resolved earlier on, and you have the chance to go back over things that you may have previously believed to be right, and re-assess them with your new knowledge gained clinically, rather than from a textbook. You also pick up valuable tips in clinical training in these feedback sessions”	BDS 4	24	Yes	10	20	18	20
“The only way that I can think dentistry has changed me is infection control. When I clean the house I do low risk areas first then move onto high risk - weird!”	BDS 5	26	No	4	20	19	11

4.6 Discussion

The present Chapter describes the usefulness of the self-reported Kember's Reflection Questionnaire as a method to assess dental students' and trainees' reflection habits. Taking into consideration that peer-assessment can encourage self-reflection (Schön, 1983; Boud and Falchikov, 2007), and that reflection is key to learning (Boud *et al.*, 1985), the formative peer-assessment and peer-feedback protocol needs to identify whether the peers' exercise is helping students to engage in reflective practice and, ideally, its depth. The results of the present Chapter show the various levels of self-reported reflective thinking habits of undergraduate dental students and postgraduate trainees following clinical and PhD postgraduate programmes. Further, it details the relationship between these reflective habits and the academic performance of undergraduate students.

The internal reliability coefficients of the RQ scores were acceptable and consistent with previous reports of the same instruments used in students from health sciences (Occupational therapy, Physiotherapy, Radiography and Nursing) (Kember *et al.*, 2000; Leung and Kember, 2003), nursing (Lethbridge *et al.*, 2011), polytechnic (Lim, 2011), medicine (Ambrose and Ker, 2014), and arts and mathematics (Phan, 2008). These results allow for the first hypothesis of this Chapter to be accepted: The results of the studied RQ are reliable.

The mean of all subjects' scores showed the RQ did differentiate between non-reflective and reflective thoughts. Thus, the non-reflective skill of Habitual Action (10.7) was significantly different ($p < 0.0001$) with both reflective skills of Reflection (17.0) and Critical Reflection (14.4). Similarly, the non-reflective thought of Understanding (17.3)

was significantly different ($p < 0.0001$) with the reflective ones of Critical Reflection (14.4) (Table 4-3). Consequently, this discernment permits to accept the second hypothesis of the present Chapter: The studied RQ scores discriminate between subjects' non-reflective and reflective thoughts.

The higher Understanding (17.3) and Reflection (17.0) mean scores indicate that students and postgraduate trainees employed these levels of thinking more than Habitual Action and Critical Reflection (Kember *et al.*, 2000; Lethbridge *et al.*, 2011; Chelliah and Arumugam, 2012). Indeed, Ashley *et al.* (2006) reported that BDS Year 4 students placed a “great deal of emphasis on practical applications of their knowledge”. Thus, consistent with a university education, they would predominantly be learning – Understanding – and integrating this new knowledge into their practice – Reflecting. On the other hand, not only would Habitual Action, characterised by its automated responses, be uncommon in busy and active learning courses, but also Critical Reflection, as it requires a major change of perspective and alteration to deep-seated beliefs (Kember *et al.*, 2000).

Reflection and Critical Reflection scores increased more than those from Habitual Action and Understanding as students progressed through the programme. As previously reported (Chelliah and Arumugam, 2012; Ambrose and Ker, 2014), this might indicate that theory and concept integration into students' and trainees' practice increases as they climb the programme ladder. Further, it is acknowledged that dental students are required to develop reflective thinking and problem solving skills as they move from well-defined problems in the classroom environment to the more uncertain and ill-defined real-life situations when they start clinical patient care (Boyd, 2008).

The undergraduate students' mean scores for Understanding, Reflection and Critical Reflection were higher, while Habitual Action ones were lower than those reported for Hong Kong undergraduate students in occupational therapy, physiotherapy, radiography and nursing (Kember *et al.*, 2000), Hong Kong health science students (Leung and Kember, 2003), Malaysian and Chinese year 3 and year 4 medical students (Chelliah and Arumugam, 2012), and those studying English language teaching in Northern Cyprus (Naghdipour and Emeagwali, 2013). Indeed, our undergraduate RQ mean scores were more comparable to those reported for Hong Kong postgraduate nursing trainees (Kember *et al.*, 2000), end-of-year 3 in practice Canadian nursing students (Lethbridge *et al.*, 2011), and year 3 and year 5 UK medical students in a Problem Based Learning

(PBL) curriculum (Ambrose and Ker, 2014). A possible explanation for this could be the particular nature of KCLDI undergraduate programme - it is a research-led degree that stimulates critical thinking (as a process of reflection) and demands extensive understanding to be successful in an intensive learning environment which promotes early responsibility for patient care (Murray *et al.*, 2003; Kings College London Dental Institute, 2013).

Clinical postgraduate and PhD trainees showed significantly higher average scores in Critical Reflection than undergraduate students (Figure 4-2). Similarly, the latter mean scores in Reflection were statistically different than those from PhD trainees. Thus, given that the levels of reflective habits from undergraduate students and postgraduate trainees are not the same, the third hypothesis of this Chapter cannot be accepted: The levels of reflective habits from undergraduate students and postgraduate trainees are the same.

These differences between students and postgraduate trainees reflective thoughts, might be explained by the professional practice model described by Schön (1983), in which unusual problems or cases, such as those possibly confronted by both clinical postgraduate and PhD trainees, might have evoked their critical reflection. Another plausible explanation might be the fact that both these groups exhibited a higher mean age at the time of the study (Table 4-2), which might indicate that they had several years of professional experience reflecting upon their professional practices (Boenink *et al.*, 2004).

The small 0.2 mean score difference exhibited by PhD trainees between Reflection (18.3) and Critical Reflection (18.1) is noteworthy; this was 1.3 for clinical postgraduates and 3.0 for undergraduate students. PhD trainees are continuously challenged by evidence and new ideas they come across and possibly have more opportunities to develop their reflection and critical reflection skills. Similar minor differences between Reflection and Critical Reflection skills have been reported by Yuen Lie Lim (2011) in a study of 111 polytechnic students in the third year of a PBL Diploma (engineering, information technology, technology for the arts, and applied sciences), where means were 15.2 for Reflection and 15.3 for Critical Reflection.

Consistent with our results, previous studies using the same RQ (Phan, 2009) and other methodologies (Aronson *et al.*, 2011; Carr and Johnson, 2013), found no gender difference in subjects' reflective habits. However, Boenink *et al.* (2004), using

reflection-evoking case vignettes, found female medical students had slightly higher reflective scores than male medical students ($p < 0.018$). In contrast, in our study the ≥ 24 years old and those with previous university degree groups showed statistically significant higher mean scores than their corresponding counterparts in their Reflection and Critical Reflection habits. This might be explained by the larger proportion of clinical postgraduate and PhD trainees in the senior and previous degree groups and the aging maturational effect (Sandars, 2009), with the concomitant development of critical thoughts and problem-solving skills distinctive of university education (Kember *et al.*, 2000; Cano, 2005).

As explained by Yuen Lie Lim (2011), the positive correlations between Understanding, Reflection and Critical Reflection scales can be explained by the questionnaire's theoretical bases that scales lie in a continuum (Lim, 2011), and contribute to an overall construct of reflective thinking (Kember *et al.*, 2000). In this context, those students and trainees with high Reflection and Critical Reflection skills may also develop their Understanding abilities, particularly in more theoretical and less practical courses (Kember *et al.*, 2000).

Reflection Questionnaire scores of Habitual Action, Understanding and Reflection from participating undergraduate students and their high stakes end-of-year examinations were positively correlated. Under these circumstances, the fourth hypothesis of this Chapter can be accepted: Undergraduate dental students' reflective habits and their high stakes academic achievements are positively correlated.

Consequently, written essays on clinical scenarios, designed to examine application, appraisal, integration and interpretation of knowledge of BDS 4 and 5 students (Paper 1b in Table 4-7), were positively correlated to students' Understanding and Reflection scores. This might indicate that essay cases with ill-defined problems required students not just to recall theoretical knowledge, but also to understand and apply it to practical situations in order to eventually solve the essay questions (Kember *et al.*, 2008). Further, "reflection is the central dynamic involved in problem solving" (Mezirow, 1991), which explains why students with higher Reflection scores in our study performed better in this kind of examination.

In contrast, short answer and short notes questions designed to assess knowledge recall of BDS 1, 2 and 3 students (Paper 1a in Table 4-7) did not correlate to any of the RQ constructs. As confirmed by Boyd (2002), the initial years of a dental curriculum have a

large number of courses and students have little time to process and learn in a meaningful way. Further, similar results were reported in a large (N=581) and detailed study on second-year undergraduate university mathematics students from the Fiji Islands (Phan, 2008), where RQ scores did not relate to their assessment marks (not reflection encouraging tutorials, written assignments and final examination, at this stage).

Single best answer and extended matching multiple choice questions (Paper 2 in Table 4-7), aimed at testing the degree of understanding of a concept or topic from different aspects of the course by knowledge recall, association and classification of concepts, were positively correlated to Understanding. Likewise, 196 second year medical students' end-of-term reflective skills correlated positively with their "cognitive" (no details of the assessment are given) academic achievements (Sobral, 2001). Further, the amount of reflective text written in a semi-structured portfolio of 91 fifth-year medical students positively correlated ($p<0.001$) with their success in a final examination comprising written case histories, multiple choice questions and essay (Lonka *et al.*, 2001).

The case of the OSCE is quite particular as it revealed a positive correlation with Habitual Action as well as with Understanding. As a competency examination and in accordance with normal practice (Schoonheim-Klein *et al.*, 2005), our students rotated between a series of 5 minute standardised stations where their understanding of a given clinical situation and their knowledge application to solve the presented clinical case was tested. To succeed at OSCE stations, students need to demonstrate that they have developed their understanding and decision making by learning to restrict themselves to the relevant features and aspects of the presented situation (Dreyfus *et al.*, 1988; Dreyfus, 2004). Our results correspond with those in a study of year 4 medical students (Carr and Johnson, 2013), though a different questionnaire did not find a positive correlation between self-reflection and the OSCE. Further, the correlation between OSCE and Habitual Action might be explained by the observation of Schön (1983) that common cases – easy or expected OSCE stations, for example – could have possibly been approached by competent students in an habitual way as they were used to the presented scenarios, performing automatically with little conscious thinking. It should be borne in mind that despite Habitual Action being a non-reflective action, this response would be appropriate, for example, in emergency circumstances where actions are automatic without conscious thoughts, after which one can reflect about these

actions. However, it should not be relied on too often, especially in “complex and challenging situations, or to advance knowledge” (Lethbridge *et al.*, 2011).

The Clinical Reasoning Examination (CRE), based on a similar assessment used in the Membership of the Joint Dental Faculties of The Royal College of Surgeons of England, was introduced at KCLDI for BDS year 5 Finals to replace the “single unseen case”. The aim is to assess communication and clinical reasoning skills in relation to clinical scenarios in the four main clinical areas of Child Dental Health, Acute Dental Care and Clinical Governance, Adult Dental Health and Clinical Diagnostic Sciences, and Oral Surgery. Thus, after 45 minutes studying four different clinical cases, students rotate between the same number of stations, each dedicated to one of the presented clinical cases, and are independently and structurally marked by two examiners. Students are marked on their level of competence in clinical reasoning of the case by articulating not only their deep knowledge and understanding, but also their ability to apply it to clinical problems (problem solving). These examination results and their positive correlation with Understanding and Reflection can again be supported by Schön's observations (1983), as unusual problems or cases – e.g. clinical reasoning scenarios – and especially those with higher implications, lead to higher order thinking.

In line with this, Dreyfus *et al.* (1988) suggest deliberation should occur before action is taken when outcome performance is crucial and time permits. Further, in a recent study (Ambrose and Ker, 2014), reflection was identified as a recurrent learning mechanism in medical students and correlated with “knowledge of actions to take”, which, coincidentally, is one of the assessment criteria of the CRE. Moreover, and according to Leung and Kember (2003), students with high Understanding and Reflection scores, as those good performers in the CRE and essays, employ a desirable deep approach to learning, which, as stated by Rolfe and McPherson (1995), is characterised by “the active pursuit of understanding and application of knowledge”, with the information being stored in the long-term memory (Harden and Laidlaw, 2013).

The Case Presentation examination is part of BDS Year 5 Finals, in which students present a clinical case they have treated in a poster format. Subsequently, two examiners check the actual treatment in the patient's mouth and interact with the student on the diagnosis and rationale of the treatment provided, seeking to reveal the student's level of understanding. It was expected that students' scores from this examination would correlate with the Understanding scale of the RQ; however, this was not the case. A

possible explanation for this might be the effect of unrestricted access to staff during the planning and treatment phase of the clinical case, meaning that students were not required to think for themselves and hence not reflecting. Another likely reason for this could be that other end-of-year academic examinations' demands occur simultaneously and might have relegated the Case Presentation to a less desirable surface approach (Cano, 2005).

Regardless of the positive correlations between Habitual Action, Understanding and Reflection scores with different examination methods, this was not seen with Critical Reflection. That is, students with high Critical Reflection scores did not perform better or worse in any end-of-year examinations. The reason for this might be found in the conceptual design of the RQ itself. Thus, the "Critical Reflection" dimension of Kember *et al.* (Kember *et al.*, 2000), based on Mezirow's "Premise Reflection" (Mezirow, 1991), are the highest levels of reflective thinking and imply the possibility for perspective transformation "guided by the identifying and judging of presuppositions" (Mezirow, 1991). This has been described as a difficult (Jonas-Dwyer *et al.*, 2013), lengthy and painful process (Kember *et al.*, 2000), associated with future intentions (Ambrose and Ker, 2014), and therefore not often observed (Lethbridge *et al.*, 2011). A longitudinal assessment programme based on a meaningful experience interpretation (Mezirow, 1991), might help to ascertain a correlation with Critical Reflection.

The results of the present Chapter contribute to five aspects regarding dental students' reflective habits that could prove useful for dental teaching and learning.

Firstly, and to the authors' best knowledge, this is the first study which reports the assessment of dental undergraduate students' and postgraduate trainees' reflective habits following the four scales of the RQ protocol developed by Kember *et al.* (Kember *et al.*, 2000). This is particularly important as this instrument has been utilised in a wide variety of disciplines which permits trustable comparisons of students' reflective habits.

Secondly, knowing students' scores in each of the RQ constructs could be helpful in investigating the effects of a given teaching and learning environment or a specific course or intervention on the reflective thinking of students.

Thirdly, some students are better reflectors than others, and thus, knowing individual student reflective patterns could be used to assist those less-able reflectors to develop

this skill and ideally improve their performance.

Fourthly, as reflective thinking did not show major distinctions between males and females, our attention in tutoring should be focused on younger students (≤ 23 years old) and those without a previous university degree, for they presented significantly lower Reflection and Critical Reflection scores.

Fifthly, a comprehensive assessment programme including a variety of assessment methods could stimulate more than a single construct of reflective thinking. Thus, whilst multiple choice questions encourage them to “understand”, ill-defined problems such as those from essays and clinical reasoning exercises would elevate this into the reflection scale.

However, the present Chapter study has some limitations. Probably the most evident is that it is only an observational study. As such, the results can only insinuate some possible roles for the Reflective Questionnaire. Thus, it would be desirable to have a study design allowing for pre- and post-observation reflection scores of similar groups during specific periods and examine the influence of different variables. Further, an intervention process could be included to examine its effect on students' reflective skills and to investigate whether the RQ is sufficiently sensitive to detect changes as a result of the intervention. Another drawback is the absence of any kind of assessment method of clinical postgraduate and PhD trainees to correlate to their reflective habits.

4.7 Conclusions

The results demonstrate that dental students and postgraduate trainees' non-reflective and reflective habits can be assessed reliably and validly using the self-reported Kember *et al.* Reflection Questionnaire (2000). Consequently, the long-standing tradition that the “collective wisdom” of the staff educators is an adequate source for evaluating students' reasoning (Hendricson *et al.*, 2006), can at least now be complemented with this reliable and valid questionnaire.

The studied Reflection Questionnaire, as a diagnostic instrument, not only allowed the determination of reflective and non-reflective skills of dental students, but also their correlation with different assessment methods. Thus, its use for a pre- and post-intervention evaluation of reflective skills, after prospectively using the peer-assessment and peer-feedback protocol, as it will be described in the next chapter, appears appropriate.

Chapter 5

Implementing a Structured Protocol for Dental Students' Prospective Peer-Assessment and Peer-Feedback

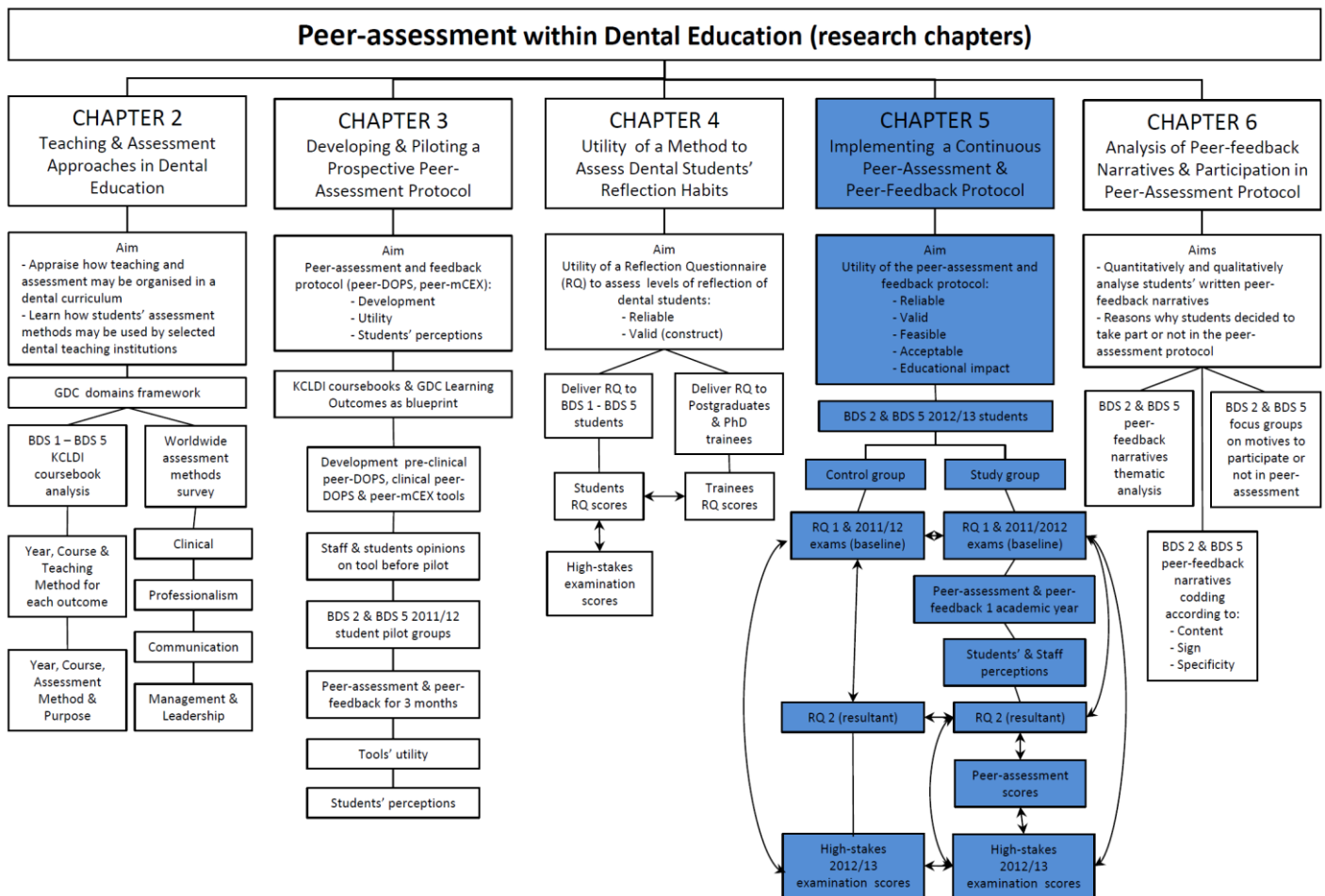


Figure 5-1 Flowchart of the study's research chapters. The fifth chapter is highlighted to help in providing an overall view.

5.1 Introduction

The clinical environment focuses students on actively learn from integrated real problems and affords teachers the opportunity to model professional thinking, behaviour and attitudes (Spencer, 2003). However, time pressures, staff competing demands and increasing numbers of students may adversely affect this (Spencer, 2003) (Chapter 1, Clinical workplace-based assessment and the role of peers in page 81). Current practice of WPBA has also been criticised for its absence of direct observation (Driessen and

Scheele, 2013), lack of informed and immediate feedback, difficulties in finding staff assessors (Quantrill and Tun, 2012) and limited opportunity for reflection and discussion (Spencer, 2003).

After the positive results of the peer-assessment protocol piloted in Chapter 3, with regards to the ability of undergraduate students to reliably peer-assess one another in all four GDC domains, and their perceived positive perceptions, it seems reasonable to think whether this protocol implemented in a longer period of time and in a larger sample of students, would help to overcome, at least partially, the current WPBA criticism described above. Probably most importantly, the protocol's potential to foster students' reflection and academic performance needs to be ascertained.

This belief has been encouraged by the recent published study of Ali *et al.* (2014), who compared the ratings of a formative peer-assessment of second year undergraduate dental student's clinical competence in simulated tooth extractions, with the ones of an equivalent, one week later, summative assessment scored by academic supervisors. Hence, in similar conditions to our protocol, students working in pairs alternated roles of assessor and trainee with their pre-clinical partner. After performing two tooth extractions in 60 minutes, the assessor acting student, following a given assessment criteria, graded and most importantly, provided verbal and written feedback to the trainee student. Despite results showing weak statistical correlations, 84% of students were awarded the same rating by peers and academic supervisors. Further, whilst a 5.7% who passed the peer-assessment later failed the summative one, a notable 10% improved their peer-assessment score in the later supervisor summative assessment. As the authors presented it, these latter results could be related to students learning from their formative peer-assessment experience.

Despite the Ali *et al.* (2014) study encouraging results for dental students' peer-assessment, together with the ones presented in Chapter 3, it seems that peer-assessment in dental education as a method to contribute to each other's learning process, is underexplored as no study has been published on the prospective use of peer-assessment and peer-feedback in dental undergraduate students, despite the potential dental peers have to contribute to each other's learning process, as described in Chapter 1 (page 53).

Furthermore and from this "learning" point of view, a repetitive peer-observation, peer-assessment and provision of a constructive and immediate dialogic peer-feedback

(agreeing challenges to change performance) (Sargeant *et al.*, 2011), might contribute to each other's learning process encouraging self-reflection (Schön, 1983; Ashley *et al.*, 2006; Boud and Falchikov, 2007; Miller *et al.*, 2010) and critical skills (Dochy *et al.*, 1999; Speyer *et al.*, 2011), and by doing so improve their clinical performance (Boursicot *et al.*, 2011).

5.2 Aim

Following a pilot study (Chapter 3), this Chapter aimed to evaluate the reliability, validity, feasibility, acceptability and educational impact of the same structured protocol of formative, prospective peer-assessment of undergraduate pre-clinical and clinical dental students' skills as a framework for the provision of immediate peer-feedback. Further, the influence of the peer-assessment protocol on the students' reflective skills was also examined.

5.3 Hypotheses

- i. The implemented structured peer-assessment protocol is reliable.
- ii. The implemented structured peer-assessment protocol is valid.
- iii. The implemented structured peer-assessment protocol is feasible.
- iv. The implemented structured peer-assessment protocol is acceptable.
- v. The implemented structured peer-assessment protocol has a positive educational impact.
- vi. The peer-assessment exercise stimulates students' reflective skills.

5.4 Materials and Methods

5.4.1 Instruments for peer-assessment and feedback

The three peer-assessment instruments developed for this research (Chapter 3, page 110), were implemented in a larger sample and for a longer period of time.

Consequently, pre-clinical work-based peer-DOPS (Appendix 12), designed for peer-

assessment of any training procedure performed at the simulation skills laboratory, along with both clinical workplace-based peer-DOPS (Appendix 13), planned for the peer-assessment of whichever clinical procedure students performed on their patients, and the peer-mCEX (Appendix 14), meant for the peer-assessment of patients' clinical examinations/assessment, were used as a framework for a continuous and structured peer-assessment and peer-feedback protocol of undergraduate pre-clinical and clinical dental students, respectively.

5.4.2 Instruments to assess academic achievement outcomes

In order to establish the possible effects of the studied peer-assessment and feedback protocol on students' academic performance, their high-stakes end-of-year examination marks from the previous 2012 academic year (baseline) as well as those from the actual 2013 studied period (resultant), were collected and subsequently correlated to the peer-assessment variables.

Two student cohorts were studied:

- a) BDS 2 baseline end-of-year examinations included written short answer questions (SAQs) and short note questions (SNQs) which constituted Paper 1, along with online multiple choice questions (MCQs), which formed Paper 2. The resultant end-of-year examinations comprised the same previous instrument formats (Paper 1 and Paper 2), supplemented by a hurdle clinical skills examination with an oral (viva) and practical components.
- b) BDS 5 baseline end-of-year examinations involved essays on clinical scenarios (Paper 1), online MCQs (Paper 2), and an Objective Structured Clinical Examination (OSCE); resultant end-of-year examinations (Finals) encompassed all three former formats with the addition of a Clinical Reasoning Examination (CRE) and a Case Presentation.

5.4.3 Instrument for reflection skills assessment

The self-reported Reflection Questionnaire (Kember *et al.*, 2000) (Chapter 4) was used to examine the influence of the peer-assessment and peer-feedback protocol on students' reflective skills, both at the beginning (baseline) and end (resultant) of the 2012/2013 academic year.

5.4.4 Implementation and data collection

One researcher (JT) supervised the whole peer-assessment process. In June and September 2012 (baseline), all 309 BDS 2 (N=154) and BDS 5 (N=155) students enrolled at KCLDI received an electronic invitation from the Director of Education to voluntarily participate in the peer-assessment and peer-feedback exercise, which started by completing a Survey Monkey® online version of the self-reported Kember RQ.

The same researcher (JT) also delivered several 15-minute tutorials to give all pre-clinical BDS 2 Conservative Dentistry and all clinical BDS 5 Primary Dental Care students a detailed explanation of the study peer-assessment protocol, emphasizing the possible benefits of feedback and specifically exposing the findings of Hattie and Timperley (2007).

Those students who consented to participate in the peer-assessment exercise, received a 45-minute training and familiarisation session (delivered by JT) on observation, peer-assessment, peer-feedback, action plan and completion of the peer-DOPS and peer-mCEX instruments. They were also encouraged to undertake a brief self-reflection on the received feedback and action plan and record it in a personal reflection diary, immediately after each peer-assessment and feedback encounter.

In September 2012 and following a naturalistic design (Crossley *et al.*, 2007; Boursicot *et al.*, 2011) in that the selection of peer observers and procedures was natural and uncontrolled, BDS 2 students attending the Conservative Dentistry skills laboratory course organised themselves in pairs to work at neighbouring phantom heads. Similarly, BDS 5 clinical partners registered for Primary Dental Care, alternated the dentist/assistant roles randomly allocated each session. In both cohorts, each student acted as “observer” and “trainee” during the first half of the day and then switched roles during the second half of the day.

BDS 2 students customarily performed their own procedures while “observing” each peer’s pre-clinical work every 15 minutes (to avoid interfering with their own pre-clinical work). BDS 5 students performed their usual clinical activities in pairs so that the assistant student “observed” the dentist student while treating the patient together.

In each case, the observed procedure was then used to formatively assess each of the peer-DOPS and peer-mCEX domains, and act as a grounded framework to provide

informed written (in the actual peer-assessment form) and verbal terminal dialogic feedback (Walsh *et al.*, 2009), as well as to agree an appropriate action plan to address any developmental needs, at the end each session. Finally, after signing the forms and placing them in a specially designed delivery box, students reflected on the feedback and action plan and noted their thoughts in a private reflection diary. In order to facilitate and stimulate the reviewing of their challenges and reflections, all feedback and action plans were uploaded (within 24 hours) to a password protected personal electronic peer-feedback and reflection diary as part of the King's e-Learning and Teaching Service (KEATS) platform (Figure 5-2).

KING'S College LONDON | [KING'S MAIN SITE](#) | [INTERNAL](#) | [KEATS HOME](#) | [JORGE TRICIO PESCE \(LOG OUT\)](#)

KEATS (KING'S E-LEARNING AND TEACHING SERVICE)

Peer-Feedback diary

These are your received feedback & agreed challenges/actions on 30/10/2012:

- Look at radiographs at the same time as clinical examination. Refer to radiographs earlier.
- Check assessment sheet to make sure haven't missed any area. e.g. oral hygiene.

These are your received feedback & agreed challenges/actions on 13/11/2012:

- Good time management and organisation of equipment.
- Getting accurate reading with apex locator. E.g. wash with sodium hypochlorite and try again and dry with paper points.

These are your received feedback & agreed challenges/actions on 15/11/2012:


- Good time management.
- Improve knowledge of protaper sequence.

These are your received feedback & agreed challenges/actions on 29/11/2012:

- Although not xxx fault as patient attended late, she worked fast and efficiently.

Peer-Feedback diary from your DOPS and Mini-CEX

Welcome to the peer-feedback blog. This has been developed for you to have access to all your DOPS and Mini-CEX peer-feedback and agreed challenges/actions forms. As presented to you earlier, it is good practice to review these past comments and reflect on them as a way to overcome any difficulty in your learning process.

 [Peer-Feedback diary from your DOPS and Mini-CEX](#)

Jump to...

Figure 5-2 Example of a BDS 5 student's peer-feedback diary designed to help them recall their peer-feedback and agreed challenges to stimulate self-reflection (the name of the student has been hidden).

Participating students freely decided when to peer-assess each other; whether to submit the completed forms in the delivery box, the number of encounters to complete; whether to stop participating and ask for the return of their completed forms.

In order to investigate students' and their tutors perceptions of the prospective peer-assessment exercise, during the last two weeks of peer-assessment they were asked to anonymously answer the following questions using a 5-point Likert-scale (Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree):

To what extent do you agree that the peer-assessment and feedback protocol used in this study:

- a) assessed areas that correspond to the activity in the pre-clinic/clinic?
- b) was easy and straightforward to use?
- c) could be introduced in the future to all students at King's College London Dental Institute as part of their pre-clinical/clinical education?
- d) was acceptable and fair?
- e) has helped you to identify learning needs and to improve your performance?

Questions b) and e) were not presented to tutors. Both student and tutor questionnaires included a final free text box for any further comments / problems / suggestions.

Lastly, before the study ended (May 2013), a third and final invitation from the Director of Education was sent asking all BDS 2 and BDS 5 students to complete the end of the academic year (resultant) RQ.

5.4.5 Students' stimulus

All students participating in the study were offered entry to a prize-draw for a Kindle Fire HD (one for each BDS year) and receive a "Research Participation Certificate" signed by the KCLDI Director of Education for their portfolio (which were delivered before the end of the academic year).

5.4.6 Data analysis

All collected data from completed peer-DOPS and peer-mCEX forms was used to evaluate the reliability, validity (content, criterion and construct), feasibility, acceptability and educational impact of the peer-assessment and feedback protocol. Furthermore, RQ scores were used to examine the influence of the peer-assessment protocol on the students' reflective skills.

As in the piloting (Chapter 3), peer-assessment scores were assigned a numerical value from 1 to 6. Thus, the “Starting to develop” initial stage of ability was given a score 1, the “Show initial capability” a score 2 and so on until the highest “Show constant extremely good ability”, which was given a score 6.

Peer-DOPS and peer-mCEX forms data completed by students were manually digitised by one researcher (JT) into a spreadsheet, whilst RQ responses were electronically downloaded from Survey Monkey® into the same combined spreadsheet. Scores from both RQs (baseline and resultant) and peer-assessment were checked for normality assumptions using histogram and box plot before carrying out any parametric analysis.

Descriptive statistics using mean, standard deviation, percentage and range were used to summarise participants' characteristics as well as the total number of observations, average of encounters per student, peer-assessed procedure categories, peer-DOPS and peer-mCEX scores and observation and feedback times. RQ traits were also analysed using descriptive statistics while its reliability was quantified by computing Cronbach's Alpha coefficient.

A G-study (Generalizability study) (Cronbach and Shavelson, 2004) was used to determine the contribution that all relevant factors (trainee, occasion, items, and their interactions) made to the results, independently for each peer-assessment form, and thus reflect their reliability (Crossley *et al.*, 2002a). Additionally, a D-study (Decision study) was used to estimate the effect of the number of peer-assessment encounters on the forms reliability. Thus, a G-study with a crossed three-facet (trainee \times occasion \times items) random-effects using items as a fix facet was used for pre-clinical peer-DOPS (assessor students were crossed within trainee students as they worked in fixed pairs), and a nested three-facet random-effects using items as a fix facet was used for clinical peer-DOPS and peer-mCEX (assessor students were most of the times nested (unique) within trainee students as they worked in random pairs).

Descriptive statistics were used to study the applicability and thus the content validity of both peer-DOPS and peer-mCEX items to be assessed by peers. Thus, the number of “unable to comment” or missing values for each item was investigated to determine whether they were assessing “valid” content. The presence of considerable “unable to comment” or missing values fields in a specific item would mean the item was not applicable. Further, peer-DOPS and peer-mCEX criterion validity was studied using a

Pearson correlation analysis between the mean peer-assessment score students were given for each of the different tools along the study period, with the mean of all their resultant high-stakes end-of-year examinations marks, which were assessing the same outcomes and therefore considered as the gold standard (Streiner and Norman, 2008).

Finally, construct validity of the peer-DOPS and peer-mCEX was assessed in four different ways by analysing:

- a) the forms internal structure using Cronbach's Alpha coefficients.
- b) the homogeneity of all three scales through inter-item correlation and item-total correlations.
- c) whether students were able to differentiate performance through one way analysis of Variance (ANOVA) among the 11 preclinical peer-DOPS items, 13 clinical peer-DOPS items, and 9 peer-mCEX items, separately.
- d) whether students could detect improvement in peers' performance over time through repeated measures ANOVA.

The feasibility of the peer-assessment protocol was studied by descriptive statistics analysing observation and feedback time reported by students in all completed peer-DOPS and peer-mCEX forms. This was complemented by students' and tutors' anonymous answers to both perception questions: "To what extent do you agree that the peer-assessment and feedback protocol used in this study was easy and straightforward to use?", and "...could be introduced in the future to all students at King's College London Dental Institute as part of their pre-clinical/clinical education?"

Further, the proportion of peer-DOPS and peer-mCEX forms delivered fully completed containing a feedback and action plan was also considered to study feasibility. Finally, following the UK Committee of Postgraduate Dental Deans & Directors (COPDEND) standards for Foundation Training dentists of one evaluation of performance per month (COPDEND, 2009), and as the present study lasted 8 months, the number of students who completed 8 or more peer-assessment forms was recorded.

Both students' and tutors' acceptability of the protocol was examined through descriptive statistics of the anonymous answers to the perception question "To what extent do you agree that the peer-assessment and feedback protocol used in this study was acceptable and fair?"

The educational impact of the peer-assessment protocol was studied through four different paths:

a) The peer-assessment and peer-feedback “effect size” on progress as described in Chapter 1 (Educational Impact in page 51) according to Hattie and Timperley (2007), was calculated for those students who exercised the peer-assessment protocol as a whole group and per the number of encounters they completed, as well as for those who did not take part in the protocol.

Thus, the difference between the resultant (2013) and the baseline (2012) (before anyone started to use the peer-DOPS and/or peer-mCEX tools), end-of-year examination marks, divided by a composite standard deviation, gave us the “effect size” (*d*) which helped to understand the impact of the intervention (peer-assessment and peer-feedback) over the academic year (Hattie, 2012).

b) As a supplement of the effect size, a paired-samples t-test was used to compare the baseline (2012) and resultant (2013) end-of-year examination marks of those students who followed the peer-assessment protocol throughout the academic year. The same comparison was also performed for those students who did not follow the protocol, as well as for the whole class.

c) An analysis using independent-samples t-test and effect size was performed comparing the students' resultant (2013) mean end-of-year examination marks, and each of their high-stakes end-of-year Finals between those students who practiced the peer-assessment and peer-feedback protocol, with those who did otherwise (control group).

d) As a result of students' anonymous answers to the perception question “To what extent do you agree that the peer-assessment and feedback protocol used in this study has helped you to identify learning needs and to improve your performance?” using descriptive statistics.

The influence of the peer-assessment and feedback protocol on the students' reflective skills was studied, initially, by comparing baseline (2012) and resultant (2013) students' RQ scores from those students who followed the peer-assessment protocol as well as from those who did not, using the effect size formula and a paired-samples t-test. Subsequently, and again calculating the effect size in addition to independent-samples t-test, the relationship between the number of peer-DOPS and/or peer-mCEX students completed with their resultant (2013) RQ scores, also was explored.

All analyses were carried out using SPSS[®] version 21, except for the generalizability coefficient which was calculated using the software EduG 6.1e.

5.5 Results

5.5.1 Participants

A total number of 157 volunteer students (102 females and 55 males, aged 20 to 42, mean=23.9, sd=3.3) participated in the study. The pre-clinical group comprised 71 BDS 2 students in total (46% of the class), 40 of whom agreed to participate in the peer-assessment exercise, while the remaining 31 BDS 2 students completed both baseline (2012) and resultant (2013) RQs but did not take part in the peer-assessment protocol. Further, the clinical group comprised 86 BDS 5 students (55% of the class), 68 of whom agreed to take part in the peer-assessment protocol, whereas the other 18 BDS 5 students completed both RQs but no peer-assessment forms.

Thus, those 108 students (40 BDS 2 and 68 BDS 5) who practiced the peer-assessment protocol became the study group, whilst those 49 (31 BDS 2 and 18 BDS 5) who only completed both RQs were considered the control group. At the end of the study, a total of 26 (65%) BDS 2 and 43 (63%) BDS 5 students from the study group, along with 9 tutors (4 pre-clinical and 5 clinical) completed the end-of-study anonymous perceptions' questionnaire of the prospective peer-assessment protocol.

5.5.2 Peer-assessment general data

Students who participated in the peer-assessment exercise completed a total of 1169 peer-DOPS and peer-mCEX forms during the whole study period. More specifically, 40 participating BDS 2 students completed 516 pre-clinical peer-DOPS forms with a range of 1 to 25 and an average of 12.9 forms each student, whilst the 68 participating BDS 5 students filled in 576 clinical peer-DOPS and 77 peer-mCEX with a range of 1 to 27, and an average of 9.6 forms each. As a consequence of both groups wide range of peer-assessment encounters, participating students were divided in three groups according to the number of peer-DOPS and peer-mCEX they completed, as shown in Table 5-1.

Table 5-1 Number of BDS 2 and BDS 5 students according to the number of peer-assessment encounters completed.

BDS 2				BDS 5		
Group	N. of students	N. of peer-DOPS	Average encounters per student	N. of students	N. of peer-DOPS & peer-mCEX	Average encounters per student
1-4 completed forms	10 (25%)	38 (7%)	3.8	28 (41.2%)	88 (13%)	3.1
5-9 completed forms	11 (27.5%)	96 (19%)	8.7	18 (26.5%)	128 (20%)	7.1
≥10 completed forms	19 (47.5%)	382 (74%)	20.9	22 (32.3%)	437 (67%)	19.8
Total	40	516	12.9	68	653	9.6

All the differently grouped pre-clinical (BDS 2) and clinical (BDS 5) procedures students assessed each other, along with their own rated case complexity for each encounter, are shown in Table 5-2 and Table 5-3 respectively.

Table 5-2 Number of pre-clinical procedures BDS 2 students peer-assessed each other using peer-DOPS according to the perceived complexity assigned to each procedure.

BDS 2 peer-DOPS					
Pre-clinical procedures	Case complexity				Total
	Low	Moderate	High	Unstated	
Amalgam	2	47	14	2	65 (12.6%)
Caries removal	5	9	3	3	20 (3.9%)
Cavity preparation	1	1	-	-	2 (0.4%)
Composite	17	145	15	12	189 (36.6%)
Core restoration	-	-	1	-	1 (0.2%)
Fissure Sealant	11	10	-	-	21 (4.1%)
Pulp cap	-	-	1	-	1 (0.2%)
Root canal treatment	11	102	36	12	161 (31.2%)
Rubber dam placement	1	3	-	1	5 (1.0%)
Temporary restoration	13	24	2	3	42 (8.1%)
Direct veneer	3	3	1	-	7 (1.4%)
Wax-up tooth morphology	2	-	-	-	2 (0.4%)
Total	66 (12.8%)	344 (66.7%)	73 (14.1%)	33 (6.4%)	516 (100%)

Table 5-3 Number of clinical procedures BDS 5 students peer-assessed each other using peer-DOPS and peer-mCEX according to the perceived complexity assigned to each procedure.

BDS 5 peer-DOPS and peer-mCEX					
Clinical procedures	Case complexity				Total
	Low	Moderate	High	Unstated	
Amalgam	8	16	3	0	27 (4.1%)
Bridge	5	8	5	0	18 (2.8%)
Composite	62	33	13	0	108 (16.5%)
Crown	33	31	39	10	113 (17.3%)
Denture	23	31	23	6	83 (12.7%)
Examination	13	13	15	5	46 (7.0%)
Inlay	0	3	3	0	6 (0.9%)
Onlay	3	8	3	0	14 (2.1%)
Root canal treatment	3	59	38	15	115 (17.6%)
Scaling and root planning	13	31	3	0	47 (7.2%)
Teeth splinting	0	3	0	0	3 (0.5%)
Treatment planning	0	15	13	3	31 (4.7%)
Veneer	0	3	3	0	6 (0.9%)
Unstated	8	24	5	0	36 (5.5%)
Total	170 (26.0%)	278 (42.6%)	166 (25.4%)	39 (6.0%)	653 (100%)

The peer scores in both groups of students ranged from 1 to 6 and were normally distributed (mean=4.1, sd=1.5, median=4.0, mode=5). The mean and standard deviations of the various peer-DOPS and peer-mCEX assessment measures from the BDS 2 and BDS 5 groups are summarised in Table 5-4.

Table 5-4 Summary statistics (mean (sd)) of various measures of BDS 2 and BDS 5 peer-assessment forms.

Variables	BDS 2 peer-DOPS	BDS 5 peer-DOPS	BDS 5 peer-mCEX
Number of observations	516	576	77
Overall observation time (in min)	116 (79.8)	85 (38.1)	56 (32.9)
Overall feedback time (in min)	8 (3.7)	5 (2.9)	5 (1.9)
Overall score (1 to 6)	3.5 (1.3)	5.2 (0.6)	5.3 (0.5)
Utility of <u>Giving Feedback</u> (1 to 6)	3.7 (1.1)	4.1 (1.1)	4.4 (1.3)
Utility of <u>Receiving Feedback</u> (1 to 6)	3.7 (1.1)	5.4 (0.8)	5.4 (1.0)

5.5.3 Reflection Questionnaires general data

Baseline (2012) RQ scores from all 157 sampled students ranged from 1 to 5 and were normally distributed (mean=3.7, sd=1.0, median=4.0, mode=4). Cronbach's Alpha internal reliability coefficient was 0.702. The adding of these baseline (2012) scores to conform each scale value, revealed a mean of 11.3 (sd=3.0) for Habitual Action, 17.5 (sd=1.7) for Understanding, 17.0 (sd=2.2) for Reflection, and 14.4 (sd=2.5) for Critical Reflection. Similarly, resultant (2013) RQ scores from the same 157 students showed the same range and comparable distribution (mean=3.8, sd=1.2, median=4.0, mode=4), whereas the Alpha coefficient was 0.731. The mean scale values this time were 11.4 (sd=3.6) for Habitual Action, 17.7 (sd=3.0) for Understanding, 17.5 (sd=2.9) for Reflection, and 14.8 (sd=3.1) for Critical Reflection.

5.5.4 Reliability of peer-DOPS and peer-mCEX

The absolute Generalizability coefficient for pre-clinical peer-DOPS forms was 0.724, while for the clinical peer-DOPS and peer-mCEX forms it was 0.787 and 0.795, respectively.

Table 5-5 D study for different number of encounters for pre-clinical peer-DOPS, clinical peer-DOPS and peer-mCEX.

Number of encounters	BDS 2 peer-DOPS	BDS 5 peer-DOPS	BDS 5 peer-mCEX
	G coefficient	G coefficient	G coefficient
1	0.563	0.601	0.632
2	0.574	0.624	0.641
3	0.592	0.631	0.655
4	0.641	0.653	0.672
5	0.667	0.661	0.689
6	0.679	0.672	0.699
7	0.690	0.687	0.721
8	0.705	0.711	0.735
9	0.719	0.746	0.747
10	0.722	0.756	0.769
11	0.728	0.765	0.775
12	0.725	0.766	0.778
13	0.727	0.769	0.789
14	0.726	0.772	0.792
15	0.724	0.787	0.795

The results of the D-study showing the G coefficient for each peer-assessment form and for the different number of encounters are shown in Table 5-5. The generally accepted G coefficient of ≥ 0.7 as sufficient for 'low-stakes' assessment situations (Beard *et al.*, 2011) was reached with 8 pre-clinical and clinical peer-DOPS, and 6 peer-mCEX forms. Additionally, the variance component estimates for the pre-clinical peer-DOPS and clinical peer-DOPS and peer-mCEX tools are shown in Table 5-6.

Table 5-6 Variance components estimates for each peer-assessment form.

Variance components	BDS 2 peer-DOPS	BDS 5 peer-DOPS	BDS 5 peer-mCEX
	Proportion of overall variance	Proportion of overall variance	Proportion of overall variance
Trainee (t)	61.8%	41.9%	45.3%
Occasion (o)	4%	0.5%	1.8%
Item (i)	3%	2.0%	4.3%
to	11.4%	30.1%	20.5%
ti	1.4%	5.0%	2.7%
oi	1%	0%	0.1%
toi	17.4%	20.5%	25.3%

5.5.5 Validity of peer-DOPS and peer-mCEX

Content Validity

There were 49 (0.9%) scores marked as "unable to comment" or missing values among 5676 scores from the 516 completed pre-clinical peer-DOPS forms. From these, 19 (0.33%) corresponded to Item 4, and 20 (0.35%) to Item 11 (Table 5-7).

Similarly, there were 80 (1.1%) scores marked as "unable to comment" among 7488 scores from the 576 completed clinical peer-DOPS. The highest corresponded to "Administers effective analgesia or safe sedation" (Item 4) with 61 (0.81%) "unable to comment" selections (Table 5-8). This is probably due to procedures not needing analgesia, such as dentures (Table 5-3). Other "unable to comment" selections were Item 7 (0.09%), Item 8 (0.08%), and Item 2 (0.05%). A much lower score was evident in the 2 (0.29%) "unable to comment" selections among all 693 scores from the 77 completed peer-mCEX, both at Item 1 (Table 5-9). Altogether, "unable to comment" or missing values accounted for 0.74% of all scores.

Table 5-7 The mean and standard deviation (sd) for the 11 peer-DOPS items for pre-clinical peer-assessment of BDS 2 students (N=516 observations).

Domain	Peer-DOPS assessment items for BDS 2	Mean (sd)
1	Knowledge of technique and dental materials used for this procedure	3.2 (1.2)
2	Preparing for procedure according to taught protocol	3.6 (1.2)
3	Technical skills, manual dexterity and instruments handling	3.4 (1.2)
4	Working position and indirect vision	3.3 (1.2)
5	Following sequence and completing accurately all steps of the procedure	3.6 (1.3)
6	Observing aseptic technique/ Infection control and safe use of instruments	3.9 (1.3)
7	Seeking help where appropriate	3.7 (1.3)
8	Managing time/punctuality effectively	3.3 (1.4)
9	Supporting and communicating effectively with colleagues and tutors	3.8 (1.2)
10	Overall ability to perform procedure	3.1 (1.2)
11	Does the trainee show insight into his/her performance?	3.2 (1.3)

Table 5-8 The mean and standard deviation (sd) for the 13 peer-DOPS items for clinical peer-assessment of BDS 5 students (N=576 observations).

Domain	Peer-DOPS assessment items for BDS 5	Mean (sd)
1	Demonstrates understanding of indications, dental materials, complications and technique of the procedure	5.3 (0.9)
2	Obtains informed consent after explaining procedure & possible complications	5.3 (0.9)
3	Demonstrates appropriate preparation pre-procedure	5.3 (1.0)
4	Administers effective analgesia or safe sedation	5.6 (0.8)
5	Demonstrates appropriate technical ability in line with usual practice	5.3 (0.9)
6	Demonstrates aseptic technique/Infection control & safe use of instruments & sharps	5.4 (0.9)
7	Deals with unexpected events or seeks help when appropriate	5.4 (0.9)
8	Completes post procedure managements	5.3 (0.9)
9	Communication skills (patient & team)	5.5 (0.8)
10	Organisation/efficiency and time management	5.3 (0.9)
11	Consideration of patient/professionalism	5.6 (0.7)
12	Overall ability to perform procedure	5.4 (0.8)
13	Does the trainee show insight into his/her performance?	5.4 (0.8)

Table 5-9 The mean and standard deviation (sd) for the 9 peer-mCEX items for clinical peer-assessment of BDS 5 students (N=77 observations).

Domain	Peer-mCEX assessment domains for BDS 5	Mean (sd)
1	Interviewing/history taking skills	5.4 (0.6)
2	Physical examination skills	5.3 (0.7)
3	Diagnostic skills and underlying knowledge base	5.1 (0.8)
4	Communication and listening skills	5.4 (0.7)
5	Clinical judgment and decision making	5.3 (0.8)
6	Consideration for patient/professionalism	5.4 (0.7)
7	Organisation/efficiency and time management	5.2 (0.7)
8	Overall clinical competence	5.5 (0.5)
9	Does the trainee show insight into his/her performance?	5.6 (0.5)

Furthermore, 81% of BDS 2 and 94% of BDS 5 students anonymously stated that they agree or strongly agree “that the peer-assessment and feedback protocol used in this study assessed areas that correspond to the activity in the pre-clinic/clinic”. Similarly, tutors’ perception of the protocol’s validity reached 88% (Table 5-10).

Table 5-10 Anonymous BDS 2 (N=26), BDS 5 (N=43) students’ and Tutors’ (N=9) perceptions (%) of the peer-assessment protocol, assessed at the end of the study.

To what extent do you agree that the peer-assessment and feedback protocol used in this study:	Source	Strongly Disagree	Disagree	Nor Agree nor disagree	Agree	Strongly Agree
...assessed areas that correspond to the activity in the pre-clinic/clinic?	BDS 2	0%	0%	19%	73%	8%
	BDS 5	0%	0%	6%	73%	21%
	Tutor	0%	0%	11%	44%	44%
...was easy and straightforward to use?	BDS 2	0%	0%	0%	50%	50%
	BDS 5	0%	3%	6%	64%	27%
...could be introduced in the future to all students at the Dental Institute as part of their pre-clinical/clinical education?	BDS 2	0%	4%	31%	62%	4%
	BDS 5	3%	3%	24%	52%	18%
	Tutor	0%	0%	0%	22%	78%
...was acceptable and fair?	BDS 2	0%	4%	11%	62%	23%
	BDS 5	0%	3%	12%	64%	21%
	Tutor	0%	0%	22%	11%	67%
...has helped you to identify learning needs and to improve your performance?	BDS 2	0%	8%	27%	58%	8%
	BDS 5	0%	9%	18%	64%	9%

Criterion Validity

The mean peer-assessment score students were given for pre-clinical or clinical peer-DOPS and peer-mCEX during the study period were correlated individually to each student's mean of all high-stakes end-of-year examination marks. Thus, BDS 2 and BDS 5 peer-DOPS scores from each participant showed positive and statistically significant correlation with their respective mean end-of-year examination mark ($r=0.583$, $p=0.0001$ for BDS 2 and $r=0.446$, $p=0.0001$ for BDS 5) (Table 5-11). However, despite peer-mCEX scores from the 17 participants students exhibited a positive correlation with their end-of-year examination mark, this was not statistically significant ($r=0.263$, $p=0.307$) (Table 5-11).

Table 5-11 Correlation between peer-assessment scores and the mean of all end-of-the-year examinations for each of the three tested instruments.

Group of students	N	Mean peer-assessment score (sd)	Mean end-of-year examination mark (sd)	Correlation	<i>p</i> value
BDS 2 peer-DOPS	40	3.5 (1.3)	65.5 (5.0)	0.583	0.0001
BDS 5 peer-DOPS	68	5.2 (0.6)	62.2 (4.5)	0.446	0.0001
BDS 5 peer-mCEX	17	5.3 (0.5)	62.9 (2.1)	0.263	0.307

Construct Validity

The internal structure of all three peer-assessment instruments was assessed using Cronbach's Alpha coefficients to indicate construct validity (Downing, 2003). Alpha was 0.964 for the 516 completed BDS 2 peer-DOPS forms, 0.960 for the 576 completed BDS 5 peer-DOPS forms, and 0.867 for the 77 completed BDS 5 peer-mCEX forms.

Further, pre-clinical peer-DOPS inter-item correlations range from a minimum of 0.571 to a maximum of 0.800. The highest item-total correlation (0.861) was for "Seeking help where appropriate" (Item 7), while Cronbach's Alpha did not improve by deleting any item and only decreased by 0.005. Regarding clinical peer-DOPS, inter-item correlations range from a minimum of 0.422 to a maximum of 0.805 and the highest item-total correlation (0.852) was for "Overall ability to perform procedure" (Item 12). Likewise BDS 2, clinical peer-DOPS Cronbach's Alpha did not improve by deleting any item. Somewhat differently, peer-mCEX inter-item correlations range from a lower minimum of 0.082 to a maximum of 0.784 and the highest item-total correlation (0.798)

was for “Overall clinical competence” (Item 8). Cronbach’s Alpha only improved to 0.881 by deleting Item 3 “Diagnostic skills and underlying knowledge”.

A third way of analysing peer-DOPS and peer-mCEX construct validity was by testing the students’ ability to detect differences in performance among the different peer-assessment domains. Thus, the scores from each of the eleven BDS 2 peer-DOPS assessed domains (Table 5-7) showed a statistically significant variance ($F=18.232$, $p<0.0001$). Although at a lower level, BDS 5 peer-DOPS thirteen domains (Table 5-8) also differed significantly ($F=2.593$, $p=0.002$), as did the peer-mCEX (Table 5-9) nine domains ($F=2.027$, $p=0.045$).

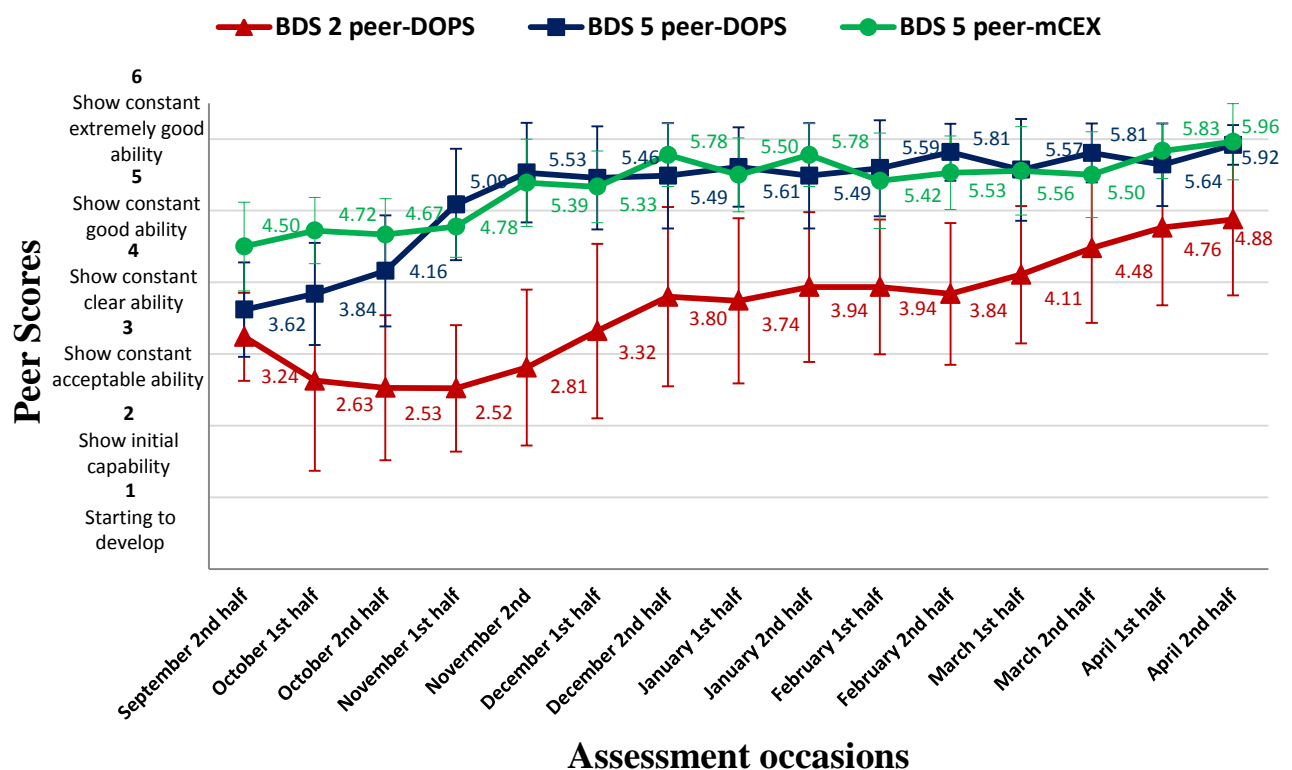


Figure 5-3 Overall mean peer-DOPS and peer-mCEX scores and standard deviations by session for BDS 2 and BDS 5 cohorts.

Students’ capability to perceive differences in performance over time using peer-DOPS and peer-mCEX was also examined. As a result, the fortnight trends in performance scores showed a progression over the academic year (Figure 5-3) which was statistically significant for BDS 2 peer-DOPS ($F=1528.652$, $p<0.0001$), as well as for BDS 5 peer-DOPS ($F=135.335$, $p<0.0001$) and peer-mCEX ($F=837.141$, $p<0.0001$). However, the

profile was uneven, especially between BDS 2 and BDS 5 scores. Thus, the former never reached the range of scores between 5 and 6 that was attained by BDS 5 peer-DOPS and peer-mCEX as early as the end of November. However, BDS 2 scores showed an almost permanent upward trend from January, which continued rising until the last encounter. Conversely, BDS 5 scores showed a steady state from November onwards.

5.5.6 Feasibility of peer-DOPS and peer-mCEX

The mean observation time (Table 5-4) for all BDS 2 peer-DOPS encounters was 116 minutes with some variability (sd=79.8). This indicates that they “observed” each other for almost the whole pre-clinical session of 180 minutes. In contrast, BDS 5 clinical students are required to see at least two patients every half day. Thus, the mean of 85 minutes in which they observed each other while performing a clinical procedure on their patients (recorded at peer-DOPS) is consistent with the 90 minute assigned sessions. The time it took for the same BDS 5 clinical students to clinically perform an examinations/assessment of a patient registered on the peer-mCEX forms (56 minutes) was less than that when treating patients. Regarding feedback time (Table 5-4), 46.7% of BDS 2 encounters were performed within the suggested 5 minutes (Norcini and Burch, 2007), while this rose to 74.9% for BDS 5 encounters. However, better evidence for this is the 100% of BDS 2 and 91% of BDS 5 students who stated they agree or strongly agree “that the peer-assessment and feedback protocol used in this study was easy and straightforward to use”. Furthermore, 66% of BDS 2 and 70% of BDS 5 students and also 100% of tutors, stated that they agree or strongly agree “that the peer-assessment and feedback protocol used in this study could be introduced in the future to all students at King’s College London Dental Institute as part of their pre-clinical/clinical education” (Table 5-10).

The total of peer-DOPS and peer-mCEX delivered fully completed with feedback and action plan reached a total of 1104/1169 (94.4%) forms. Despite being fully completed in all other items, a total of 40 (3.4%) forms were submitted in the delivery box without written feedback comments, while 25 (2.1%) did not contain transcribed agreed action plans. Finally, the number of students who completed 8 or more peer-assessment forms (at least one per month)(COPDEND, 2006) in BDS 2 was 27/40 (67.5%) whereas in BDS 5 this number was 29/68 (42.6%). Overall, these results indicate the studied protocol is feasible to be implemented in dental undergraduate pre-clinical and clinical

courses and therefore the third hypothesis of the Chapter can be accepted: The implemented structured peer-assessment protocol is feasible.

5.5.7 Acceptability of peer-DOPS and peer-mCEX

Students who anonymously stated that they agree or strongly agree “that the peer-assessment and feedback protocol used in this study was acceptable and fair” reached 85% in both BDS 2 and BDS 5. Likewise, tutors' perception of the protocol's acceptability and fairness amounted to 78% (Table 5-10). Selected BDS 2 and BDS 5 students' free text feedback comments extracted from their perceptions questionnaires are shown in Table 5-12.

Table 5-12 Selected anonymous students' and tutors' free text comments on the peer-assessment protocol.

Student / Tutor	Number of encounters	Feedback	Acceptability score
BDS 2	>13	From my experience of this process I believe giving the students a structured method to review/analyse their work has been extremely useful.	Strongly agree
BDS 2	10-13	Think that these forms should be filled out by tutors plus observing student.	Agree
BDS 2	10-13	It is dull and time consuming. Clinical partner can just tell me.	Nor Agree nor disagree
BDS 2	>13	I always felt rushed. Main problem was having enough time to fill in the forms at the end of each session. There should be time set aside for it.	Agree
BDS 2	6-9	The feedback was fairly useful.	Agree
BDS 2	10-13	More concise form. Useful but does take too long to complete.	Nor Agree nor disagree
BDS 5	10-13	Reflection is very important in clinical improvement. IPad/tablet form would be easier to track and feedback on.	Strongly agree
BDS 5	6-9	These forms should be introduced on all the clinics. Very useful and helpful for us. Students are always asking for feedback on performance.	Strongly agree
BDS 5	5-3	I feel that this exercise would be more useful for BDS 1-3 students	Agree
BDS 5	2-5	I already reflect on what went well / what could be improved after each session but don't write it down.	Agree
Tutor	-	The main reason for the success of this very important project lies with the consistent attendance of the investigator who directed his research in a non-intrusive manner and ignited the interests of all who took part.	Strongly agree
Tutor	-	Yes, they all need feedback on results	Agree
Tutor	-	Peer review and reflective conversation resulted in self-criticism and improvement of personal and clinical skills. It points out problems that they may not be aware of. Strongly recommend it.	Strongly agree

5.5.8 Educational impact

The first and second paths to assess the peer-assessment and peer-feedback protocol's educational impact analysed the effect size and the statistical significance of the differences between students' baseline (2012) end-of-year examination marks (before the use peer-DOPS and/or peer-mCEX tools), and the resultant (2013) end-of-year examination marks. This was performed for BDS 2 students (Table 5-13) comparing the mean of all high-stakes examinations as well as for those specific tests with the same format at both times, namely written SAQs (short answer questions) and SNQs (short note questions), and online MCQs (multiple choice questions). An initial comparison between baseline scores of all BDS 2 studied groups using one way ANOVA test, showed the observable small differences between mean scores were not statistically significant for the average of all examination ($F=0.232$, $p=0.948$), written SAQs and SNQs ($F=1.584$, $p=0.165$), and online MCQs ($F=1.928$, $p=0.090$). This indicates all studied groups started the intervention in similar conditions.

Subsequently, the paired t-test comparing BDS 2 baseline and resultant mean scores for all groups showed no statistically significant differences, despite an observable tendency for those 19 students who completed 10 or more peer-assessment and peer-feedback encounters to show higher resultant scores (≥ 3.3 points) (Table 5-13). In line with this, the effect size (d) on these 19 students was higher than the "gold standard" $d=0.40$ average schooling effect (Hattie and Timperley, 2007; Norcini and Burch, 2007), in all three cases: average of all examination ($d=0.59$), written SAQs and SNQs ($d=0.67$), and online MCQs ($d=0.42$). Further, those 114 students from the same class who did not take part in the peer-assessment and peer-feedback protocol, showed no statistically significant differences and small or even negative effect sizes (Table 5-13).

The same analyses was performed for BDS 5 students comparing the effect size and statistical significance of the differences between baseline and resultant means of all high-stakes end-of-year examination marks along with those specific tests using the same format at baseline and resultant observation times (Table 5-14). These were, essay on clinical scenarios, online SAQs, SNQs and MCQs, and OSCE. As with BDS 2 results, the comparison between baseline scores from all BDS 5 groups using one way ANOVA test, did not reveal statistically significant differences for the average of all examination ($F=0.783$, $p=0.562$), essay on clinical scenarios ($F=1.051$, $p=0.388$), online SAQs, SNQs and MCQs ($F=0.838$, $p=0.524$), and OSCE ($F=1.648$, $p=0.147$).

However and in contrast to BDS 2, the paired t-test comparing BDS 5 baseline and resultant mean scores showed some statistically significant differences for some groups (Table 5-14). Thus, the difference between baseline and resultant average of all examinations from those 22 students who completed 10 or more peer-assessment and peer-feedback encounters was not only statistically significant ($p=0.009$) but with high positive effect size ($d=0.74$). Even more, the effect size for those 18 students who completed between 5 and 9 encounters was also higher than the 0.40 average ($d=0.46$), though the paired t-test showed no significance.

As depicted in the same Table 5-14, it seems that the peer-assessment protocol did not have any especial effect on students' essay and OSCE marks, as all groups significantly increased and decreased their scores, respectively. However, this might also be due to other issues like for example, examination difficulty. The online SAQs, SNQa and MCQs only showed a small size effect ($d=0.33$) between baseline and resultant observations, though not statistically significant ($p=0.279$).

The third path to study the protocol's educational impact studied the effect size and statistical significance (independent-samples t-test) of the differences between resultant (2013) end-of-year examinations from those students who practice the peer-assessment protocol against those who did not. Accordingly, BDS 2 students (Table 5-15) showed that the difference in the average of all examinations between those 40 students who practiced the peer-assessment protocol (mean=65.5, sd=5.0) and those 114 who did not (mean=63.1, sd=4.9) was statistically significant ($p=0.008$) as well as with a positive effect size ($d=0.48$). More specifically, this was led by a powerful effect size of $d=0.80$ and highly significant ($p=0.001$) difference between the group of those 19 students who peer-assessed each other 10 or more times with a mean of 67.1 (sd=3.8) in their average of all examinations, and those 114 students who did not peer-assess one another (mean=63.1, sd=4.9). Furthermore, the differences between these two group scores were statistically significant and with a high positive effect size in their online MCQs ($p=0.047$, $d=0.49$), Clinical Skills Examination viva ($p=0.050$, $d=0.48$) and practical ($p=0.021$, $d=0.56$). However, there was no measurable effect on written SAQs and SNQs for any of the studied groups.

A similar outline was shown by BDS 5 students' results when comparing their Finals scores (Table 5-16). Thus, those 68 students who practiced the peer-assessment protocol showed a statistically significant difference ($p=0.028$) and a positive effect size ($d=0.36$)

in their average of all examination (mean=62.2, sd=4.5) when compared to those 87 students of the same class who did not take part in the peer-assessment study (mean=60.7, sd=3.7).

Once again, those 22 students' scores who peer-assessed each other 10 or more times exhibited highly statistically significant differences and powerful positive effect sizes when compared to the 87 students who did not participate in the peer-assessment protocol, in their average of all examination ($p=0.0001$, $d=0.74$), essays on clinical scenarios ($p=0.0001$, $d=0.89$), online MCQs ($p=0.0004$, $d=0.80$), and Clinical Reasoning Examination ($p=0.011$, $d=0.59$). However, this impact was not observed comparing their OSCEs ($p=0.299$, $d=0.24$) and Case Presentation's ($p=0.272$, $d=0.25$) scores as they obtained similar results.

Finally, a majority of 66% of BDS 2 and 73% BDS 5 students agreed or strongly agreed that the peer-assessment and peer-feedback protocol used in this study helped them to identify learning needs and to improve their performance (Table 5-10).

Table 5-13 BDS 2 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (*p*) (paired-samples *t*-test), and effect size (*d*), for the students' average of all end-of-year examination marks as well as specifically for those examinations with the same format both at baseline and resultant observations, according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed. The latter examinations were written SAQs (short answer questions) and SNQs (short note questions), and online MCQs (multiple choice questions).

BDS 2 baseline versus resultant end-of-year examination marks											
Studied groups		N	Average all examinations			Written SAQs, SNQs			Online MCQs		
			Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>
				<i>baseline vs. resultant</i>			<i>baseline vs. resultant</i>			<i>baseline vs. resultant</i>	
All class (baseline)		154	63.5 (9.7)	0.799	0.03	60.8 (9.3)	0.140	0.18	66.2 (10.6)	0.597	-0.06
All class (resultant)			63.7 (5.1)			62.4 (8.1)			65.6 (8.2)		
Peer-assessment	All participating students (baseline)	40	62.5 (9.3)	0.087	0.41	59.8 (9.0)	0.088	0.41	65.2 (10.1)	0.187	0.32
	All participating students (resultant)		65.5 (5.0)			62.8 (5.5)			68.1 (7.9)		
	1-4 completed forms (baseline)	10	64.6 (8.8)	0.805	-0.13	60.3 (9.1)	0.693	0.20	66.2 (11.3)	0.762	0.14
	1-4 completed forms (resultant)		63.8 (5.4)			62.1 (9.2)			67.7 (8.8)		
	5-9 completed forms (baseline)	11	61.3 (10.8)	0.430	0.37	60.5 (5.0)	0.383	0.39	63.3 (12.4)	0.551	0.36
	5-9 completed forms (resultant)		64.4 (6.2)			62.4 (4.5)			67.4 (10.1)		
	≥10 completed forms (baseline)	19	63.4 (8.3)	0.094	0.59	59.1 (9.7)	0.079	0.67	65.4 (9.3)	0.188	0.42
	≥10 completed forms (resultant)		67.1 (3.8)			63.5 (3.3)			68.7 (6.2)		
No peer-assessment	All non-participating students (baseline)	114	63.9 (9.8)	0.478	-0.10	61.1 (9.5)	0.415	0.12	66.6 (10.8)	0.168	-0.20
	All non-participating students (resultant)		63.1 (4.9)			62.2 (8.9)			64.7 (8.2)		

Table 5-14 BDS 5 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (*p*) (paired-samples *t*-test), and effect size (*d*), for the students' average of all end-of-year examination marks as well as specifically for those examinations with the same format both at baseline and resultant observations, according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed. The latter examinations were essays on clinical scenarios, online SAQs (short answer questions), SNQs (short note questions) and online MCQs (multiple choice questions), and the OSCE (Objective Structured Clinical Examination).

BDS 5 baseline versus resultant end-of-year examination marks														
Studied groups		N	Average all examinations			Essay on clinical scenarios			Online SAQs, SNQs, MCQs			OSCE		
			Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>
				<i>baseline vs. resultant</i>	<i>baseline vs. resultant</i>		<i>baseline vs. resultant</i>	<i>baseline vs. resultant</i>						
All class (baseline)		155	61.3 (4.7)	0.878	0.02	57.7 (5.0)	0.0001	0.69	63.1 (6.3)	0.744	-0.04	61.2 (6.3)	0.0001	-0.89
All class (resultant)			61.3 (4.2)			61.3 (5.0)			62.8 (5.3)			56.6 (4.0)		
Peer-assessment	All participating students (baseline)	68	60.8 (4.1)	0.059	0.33	57.1 (4.5)	0.0001	0.94	64.0 (6.2)	0.928	-0.02	62.4 (6.6)	0.0001	-0.98
	All participating students (resultant)		62.2 (4.5)			62.2 (5.2)			63.9 (5.4)			56.9 (4.3)		
	1-4 completed forms (baseline)	28	60.0 (4.1)	0.945	-0.02	56.1 (3.0)	0.002	0.80	63.5 (5.8)	0.312	-0.28	60.8 (6.5)	0.0006	-1.03
	1-4 completed forms (resultant)		59.9 (3.9)			59.9 (5.4)			61.9 (6.1)			55.6 (3.7)		
	5-9 completed forms (baseline)	18	60.5 (4.6)	0.232	0.46	57.3 (2.9)	0.0007	1.19	64.3 (6.6)	0.921	-0.04	62.1 (6.6)	0.009	-1.08
	5-9 completed forms (resultant)		62.4 (3.3)			61.6 (2.9)			64.1 (4.6)			56.2 (4.3)		
	≥10 completed forms (baseline)	22	61.6 (4.1)	0.009	0.74	58.0 (6.2)	0.0001	1.10	64.5 (6.5)	0.279	0.33	64.0 (6.5)	0.0005	-1.12
	≥10 completed forms (resultant)		64.9 (4.0)			65.2 (4.7)			66.2 (3.9)			58.1 (4.1)		
No peer-assessment	All non-participating students (baseline)	87	60.5 (4.7)	0.728	0.05	58.3 (5.3)	0.002	0.47	62.3 (6.4)	0.753	-0.05	60.3 (6.0)	0.0001	-0.63
	All non-participating students (resultant)		60.7 (3.7)			60.7 (4.7)			62.1 (5.1)			57.0 (4.5)		

Table 5-15 BDS 2 resultant (2013) mean scores, standard deviations (*sd*), statistical significance of the difference (*p*) (independent-samples *t*-test), and effect size (*d*) when comparing all groups to the 114 non-participating students (No peer-assessment), for each of the students' end-of-year examination marks (SAQs (short answer questions), SNQs (short note questions), MCQ (multiple choice questions) according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.

BDS 2 end-of-year examination marks 2013																	
Studied groups		N	Average all examinations			Written SAQs, SNQs			Online MCQs			Clinical Skills Examination viva			Clinical Skills Examination practical		
			Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>D</i>	Mean (sd)	<i>p</i>	<i>d</i>
				<i>diff. with no peer-assessment group</i>	<i>diff. with no peer-assessment group</i>		<i>diff. with no peer-assessment group</i>	<i>diff. with no peer-assessment group</i>		<i>diff. with no peer-assessment group</i>	<i>diff. with no peer-assessment group</i>		<i>diff. with no peer-assessment group</i>				
All class		154	63.7 (5.1)	0.297	0.13	62.4 (8.1)	0.872	0.02	65.6 (8.2)	0.382	0.11	64.1 (10.8)	0.517	0.08	62.1 (11.1)	0.507	0.08
Peer-assessment	All participating students	40	65.5 (5.0)	0.008	0.48	62.8 (5.5)	0.675	0.08	68.1 (7.9)	0.027	0.41	66.5 (10.9)	0.098	0.30	64.7 (9.5)	0.084	0.32
	1-4 completed forms	10	63.8 (5.4)	0.674	0.14	62.1 (9.2)	0.976	0.01	67.7 (8.8)	0.277	0.36	63.9 (9.0)	0.838	0.07	61.4 (9.3)	0.950	0.02
	5-9 completed forms	11	64.4 (6.2)	0.399	0.27	62.4 (4.5)	0.950	0.02	67.4 (10.1)	0.321	0.31	65.5 (13.0)	0.510	0.21	62.5 (8.3)	0.699	0.12
	≥10 completed forms	19	67.1 (3.8)	0.001	0.80	63.5 (3.3)	0.538	0.15	68.7 (6.2)	0.047	0.49	68.4 (10.8)	0.050	0.48	67.7 (9.8)	0.021	0.56
No peer-assessment	All non-participating students	114	63.1 (4.9)	-	-	62.2 (8.9)	-	-	64.7 (8.2)	-	-	63.2 (10.6)	-	-	61.2 (11.5)	-	-

Table 5-16 BDS 5 resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (independent-samples t-test), and effect size (d) when comparing all groups to the 87 non-participating students (No peer-assessment), for each of the students' end-of-year examination marks (SAQs=short answer questions, SNQs=short note questions, MCQs=multiple choice questions, OSCE=Objective Structured Clinical Examination) according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.

BDS 5 end-of-year examination marks 2013																				
Studied groups		N	Average all examinations			Essay on clinical scenarios			OnlineMCQs			OSCE			Clinical Reasoning Examination			Case Presentation		
			Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>
				<i>diff. with no peer-assessment group</i>			<i>diff. with no peer-assessment group</i>			<i>diff. with no peer-assessment group</i>			<i>diff. with no peer-assessment group</i>			<i>diff. with no peer-assessment group</i>			<i>diff. with no peer-assessment group</i>	
All class		155	61.3 (4.2)	0.279	0.15	61.3 (5.0)	0.326	0.13	62.8 (5.3)	0.250	0.16	56.6 (4.0)	0.463	-0.10	64.3 (8.5)	0.432	0.11	61.7 (8.4)	0.424	0.11
Peer-assessment	All participating students	68	62.2 (4.5)	0.028	0.36	62.2 (5.2)	0.066	0.30	63.9 (5.4)	0.030	0.36	56.9 (4.3)	0.873	-0.03	65.6 (8.9)	0.152	0.24	63.0 (8.5)	0.106	0.27
	1-4 completed forms	28	59.9 (3.9)	0.319	-0.22	59.9 (5.4)	0.436	-0.17	61.9 (6.1)	0.887	-0.03	55.6 (3.7)	0.125	-0.34	62.4 (6.9)	0.436	-0.17	60.6 (8.0)	0.890	-0.03
	5-9 completed forms	18	62.4 (3.3)	0.128	0.45	61.6 (2.9)	0.502	0.19	64.1 (4.6)	0.172	0.40	56.2 (4.3)	0.530	-0.18	66.4 (7.0)	0.207	0.37	65.0 (8.3)	0.079	0.50
	≥10 completed forms	22	64.9 (4.0)	0.0001	0.74	65.2 (4.7)	0.0001	0.89	66.2 (3.9)	0.0004	0.80	58.1 (4.1)	0.299	0.24	68.3 (8.3)	0.011	0.59	62.9 (8.6)	0.272	0.25
No peer-assessment	All non-participating students	87	60.7 (3.7)	-	-	60.7 (4.7)	-	-	62.1 (5.1)	-	-	57.0 (4.5)	-	-	63.7 (7.6)	-	-	60.8 (8.2)	-	-

5.5.9 Peer-assessment influence on students' reflective skills

The influence of the peer-assessment protocol was also assessed by studying the relationship between baseline (2012) and resultant (2013) students' reflection skills scores from those students who followed the peer-assessment protocol and those who did not. As described in Chapter 4, once completing the RQ, each participant received a score ranging from a minimum of 4 to a maximum of 20 for each Habitual Action, Understanding, Reflection and Critical Reflection scale. Consequently, the higher the score, the more agreement with engaging in the particular dimension that each scale measures. All groups showed higher scores for Understanding and Reflection scales both before (baseline) and after (resultant) the peer-assessment exercise, indicating that students employed these levels of thinking more than Habitual Action and Critical Reflection (Kember *et al.*, 2000; Lethbridge *et al.*, 2011; Chelliah and Arumugam, 2012). It is worth noting that the RQ assessment was identical on both occasions and hence the baseline and resultant scores therefore evaluated the real progress or regression in each of its four scales.

For BDS 2, statistically significant higher scores and positive effect sizes between baseline and resultant observations (paired t-test) were perceived in those 40 students who performed the peer-assessment exercise in Understanding ($p=0.003$, $d=0.65$) and Reflection ($p=0.004$, $d=0.63$) scales (Table 5-17). Once more, those 19 students who completed 10 or more peer-assessment encounters showed statistically significant differences and positive effect sizes in their baseline and resultant Understanding ($p=0.034$, $d=0.82$) and Reflection ($p=0.008$, $d=0.95$) scores. Additionally, and though not statistically significant, those students who completed between 5 and 9 forms also demonstrated a high effect size in Understanding ($d=0.77$) and Reflection ($d=0.52$).

In contrast, BDS 5 student's results (Table 5-18) showed a statistically significant increase in their Critical Reflection skills from baseline to resultant ($p=0.003$) only for those 22 students who completed 10 or more peer-DOPS and/or peer-mCEX encounters. The same group also demonstrated a positive effect size in their Critical Reflection skills ($d=1.04$) as did those 18 students with 5 and 9 encounters ($d=0.46$). A strong negative effect size was evident in the Habitual Action skills of those students with 10 or more peer-assessment encounters ($d= -0.65$).

The BDS 2 independent-samples t-test and effect size of the differences between

resultant RQ scores from those students who exercised the peer-assessment and peer-feedback protocol, and those who completed the RQ but did not take part in the peer-assessment protocol, showed statistically significant results and positive effect sizes in Understanding and Reflection scales (Table 5-19). Thus, those 11 students with 5 to 9 peer encounters showed statistically significant higher ($p=0.047$) Understanding scores and a positive effect size ($d=0.68$), when compared to those 31 students who did not peer-assessed each other. Similarly, those 19 students with 10 or more encounters also showed significantly higher ($p=0.011$) Understanding scores and positive effect size ($d=0.71$) than those 31 who did none. Interestingly, the difference in Habitual Action between the 40 students who did participate in the peer-assessment protocol and the 31 who did not, revealed a high negative effect size ($d= -0.46$, $p=0.05$).

Finally, BDS 5 students who completed 10 or more peer-assessment encounters ($N=22$) demonstrated a statistically significant higher Critical Reflection score ($p=0.005$) as compared to the group of 18 students who did not peer-assessed one another (Table 5-20). This difference also showed a large effect size of $d=1.41$. However, the same groups displayed a non-statistically significant difference in their Reflection score ($p=0.826$), despite had a high effect size ($d=0.78$).

Table 5-17 BDS 2 baseline (2012) and resultant (2013) mean scores, standard deviations (*sd*), statistical significance of the difference (*p*) (paired-samples *t*-test), and the effect size (*d*) for each of the four scales of the RQ according to participants' academic year and whether they participated in the peer-assessment exercise.

BDS 2 baseline versus resultant Reflection Questionnaire scores														
Studied groups		N	Habitual Action			Understanding			Reflection			Critical Reflection		
			Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>
				<i>baseline vs. resultant</i>			<i>baseline vs resultant</i>			<i>baseline vs. resultant</i>			<i>baseline vs. resultant</i>	
All who completed the RQ (baseline)		71	9.9 (2.3)	0.103	0.27	17.8 (1.2)	0.072	0.30	16.4 (2.1)	0.041	0.36	13.1 (2.3)	0.741	0.06
All who completed the RQ (resultant)			10.6 (2.8)			18.2 (1.6)			17.1 (2.5)			13.3 (2.7)		
Peer-assessment	With RQ (baseline)	40	10.0 (2.7)	0.759	-0.05	17.8 (1.3)	0.003	0.65	16.4 (1.9)	0.004	0.63	13.2 (2.6)	0.727	0.08
	With RQ (resultant)		9.9 (2.4)			18.7 (1.2)			17.7 (2.1)			13.4 (2.5)		
	1-4 completed forms (baseline)	10	10.9 (2.9)	0.643	-0.22	17.8 (1.0)	0.836	0.10	16.1 (1.4)	1	0.00	13.0 (3.1)	0.635	0.03
	1-4 completed forms (resultant)		10.4 (1.7)			17.9 (1.1)			16.1 (2.6)			13.1 (2.1)		
	5-9 completed forms (baseline)	11	9.6 (2.7)	0.794	0.12	17.9 (1.4)	0.051	0.77	16.2 (2.0)	0.235	0.52	12.5 (2.1)	0.692	0.17
	5-9 completed forms (resultant)		9.9 (2.2)			18.9 (1.0)			17.1 (1.5)			12.8 (2.1)		
	≥10 completed forms (baseline)	19	9.9 (2.4)	0.774	-0.11	17.9 (1.4)	0.034	0.82	17.0 (2.4)	0.008	0.95	13.7 (2.6)	0.829	0.07
	≥10 completed forms (resultant)		9.6 (2.8)			18.9 (1.2)			18.9 (1.2)			13.9 (1.8)		
No peer-assessment	With RQ (baseline)	31	10.2 (1.7)	0.109	0.40	17.8 (1.2)	0.810	-0.06	16.5 (1.2)	0.395	-0.02	13.0 (2.0)	0.921	0.03
	With RQ (resultant)		11.1 (2.5)			17.7 (1.9)			16.4 (2.8)			13.1 (3.0)		

Table 5-18 BDS 5 baseline (2012) and resultant (2013) mean scores, standard deviations (*sd*), statistical significance of the difference (*p*) (paired-samples *t*-test), and the effect size (*d*) for each of the four scales of the RQ according to participants' academic year and whether they participated in the peer-assessment exercise.

BDS 5 baseline versus resultant Reflection Questionnaire scores														
Studied groups		N	Habitual Action			Understanding			Reflection			Critical Reflection		
			Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>
				<i>baseline vs. resultant</i>			<i>baseline vs resultant</i>			<i>baseline vs. resultant</i>			<i>baseline vs. resultant</i>	
All who completed the RQ (baseline)		86	12.4 (1.4)	0.695	-0.20	17.2 (2.2)	0.786	0.08	17.5 (1.3)	0.865	0.17	15.5 (1.7)	0.430	0.32
All who completed the RQ (resultant)			12.1 (1.8)			17.4 (2.4)			17.8 (2.1)			16.0 (1.7)		
Peer- assessment	With RQ (baseline)	68	12.7 (2.9)	0.639	-0.20	17.2 (2.8)	0.733	0.11	18.1 (1.4)	0.945	0.15	15.6 (2.0)	0.514	0.25
	With RQ (resultant)		12.2 (2.3)			17.4 (2.3)			18.2 (2.1)			16.1 (1.8)		
	1-4 completed forms (baseline)	28	12.3 (2.5)	0.746	0.21	16.8 (3.0)	0.367	0.05	17.8 (1.5)	0.502	0.04	14.0 (3.5)	0.225	0.15
	1-4 completed forms (resultant)		12.9 (2.6)			16.9 (1.9)			17.5 (1.8)			14.5 (2.3)		
	5-9 completed forms (baseline)	18	11.7 (1.6)	0.985	0.02	16.5 (3.9)	0.274	0.18	16.9 (1.8)	0.082	0.37	14.3 (1.2)	0.603	0.46
	5-9 completed forms (resultant)		11.7 (2.2)			17.6 (1.8)			18.1 (1.7)			15.3 (3.0)		
	≥10 completed forms (baseline)	22	13.6 (2.9)	0.434	-0.65	18.1 (1.2)	0.940	0.10	18.8 (0.8)	0.492	0.25	16.4 (1.1)	0.003	1.04
	≥10 completed forms (resultant)		11.9 (2.1)			18.3 (1.6)			19.2 (2.6)			18.0 (1.9)		
No peer- assessment	With RQ (baseline)	18	12.0 (1.5)	0.659	-0.24	16.4 (2.3)	0.624	0.27	17.6 (1.0)	0.816	0.13	15.4 (1.4)	0.899	0.00
	With RQ (resultant)		11.6 (2.0)			17.1 (2.9)			17.7 (1.2)			15.4 (1.7)		

Table 5-19 BDS 2 resultant (2013) mean scores, standard deviations (sd), statistical significance (independent-samples t-test) of the difference (p), and the effect size (d), when compared to the 31 non-participating students (No peer-assessment), for each of the four scales of the RQ according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.

BDS 2 Reflection Questionnaire scores 2013														
Studied groups		N	Habitual Action			Understanding			Reflection			Critical Reflection		
			Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>
				<i>diff. with no peer-assessment group</i>			<i>diff. with no peer-assessment group</i>			<i>diff. with no peer-assessment group</i>			<i>diff. with no peer-assessment group</i>	
All who completed the RQ		71	10.6 (2.8)	0.445	-0.16	18.2 (1.6)	0.126	0.33	17.1 (2.5)	0.199	0.32	13.3 (2.7)	0.800	0.06
Peer-assessment	All participating students	40	9.9 (2.4)	0.050	-0.46	18.7 (1.2)	0.008	0.62	17.7 (2.1)	0.030	0.58	13.4 (2.5)	0.679	0.08
	1-4 completed forms	10	10.4 (1.7)	0.444	-0.28	17.9 (1.1)	0.726	0.13	16.1 (2.6)	0.753	-0.13	13.1 (2.1)	0.980	-0.01
	5-9 completed forms	11	9.9 (2.2)	0.187	-0.46	18.9 (1.0)	0.047	0.68	17.1 (1.5)	0.456	0.32	12.8 (2.1)	0.754	-0.11
	≥10 completed forms	19	9.6 (2.8)	0.070	-0.52	18.9 (1.2)	0.011	0.71	18.9 (1.2)	0.0007	1.12	13.9 (1.8)	0.319	0.30
No peer-assessment	All non-participating students	31	11.1 (2.5)	-	-	17.7 (1.9)	-	-	16.4 (2.8)	-	-	13.1 (3.0)	-	-

Table 5-20 BDS 5 resultant (2013) mean scores, standard deviations (sd), statistical significance (independent-samples t-test) of the difference (p), and the effect size (d), when compared to the 18 non-participating students (No peer-assessment), for each of the four scales of the RQ according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.

BDS 5 Reflection Questionnaire scores 2013														
Studied groups		N	Habitual Action			Understanding			Reflection			Critical Reflection		
			Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>P</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>	Mean (sd)	<i>p</i>	<i>d</i>
				diff. with no peer-assessment group			diff. with no peer-assessment group			diff. with no peer-assessment group			diff. with no peer-assessment group	
All who completed the RQ		86	12.1 (1.8)	0.659	0.29	17.4 (2.4)	0.798	0.10	17.8 (2.1)	0.915	0.05	16.0 (1.7)	0.562	0.36
Peer-assessment	All participating students	68	12.2 (2.3)	0.627	0.29	17.4 (2.3)	0.764	0.11	18.2 (2.1)	0.714	0.27	16.1 (1.8)	0.521	0.31
	1-4 completed forms	28	12.9 (2.6)	0.252	0.57	16.9 (1.9)	0.845	-0.09	17.5 (1.8)	0.748	-0.16	14.5 (2.3)	0.342	-0.48
	5-9 completed forms	18	11.7 (2.2)	0.920	0.06	17.6 (1.8)	0.699	0.19	18.1 (1.7)	0.452	0.26	15.3 (3.0)	0.920	-0.05
	≥10 completed forms	22	11.9 (2.1)	0.823	0.18	18.3 (1.6)	0.230	0.49	19.2 (2.6)	0.014	0.78	18.0 (1.9)	0.005	1.41
No peer-assessment	All non-participating students	18	11.6 (2.0)	-	-	17.1 (2.9)	-	-	17.7 (1.2)	-	-	15.4 (1.7)	-	-

5.6 Discussion

This Chapter reports the utility results of the implementation of the previously piloted (Chapter 3) prospective, formative, structured peer-assessment protocol of undergraduate dental students' pre-clinical competence and clinical performance, used as a framework for immediate dialogic peer-feedback. Further, and bearing in mind that work-based performance is case-specific (Norcini, 2003a), the protocol was designed to be implemented on a continuous basis so that several encounters could help to produce reliable results. Accordingly, BDS 2 and BDS 5 students practiced, through direct and permanent observation, how to select good evidence (Biggs and Tang, 2011) to make judgements of their peers' process of simulation exercises (pre-clinical students) and patient care (clinical students), to provide accurate peer-feedback, during the whole academic year. As highlighted previously (Chapter 3, page 129), it is important to remember that students were not asked to judge their peer's competence or performance "quality", nor the practice "volume", which might prove difficult for junior students (Norcini, 2003a), but rather their "increasing ability over time". However, it was inevitable, but desirable from an educational viewpoint, that the peer-assessment exercise fostered students to self-reflect on the "quality" of their own work (Biggs and Tang, 2011 p. 196).

Consistent with Miller's concept of work-based assessment (Miller, 1990), in order to avoid an artificial testing situation like in skill-based assessments (Wass, 2011), the design of both peer-DOPS and peer-mCEX forms required a "routine" observation of the actual job activities of students in the cognitive and pre-clinical/clinical skills, communication, professionalism and management domains (Chapter 3, Table 3-1, Table 3-2, Table 3-3, in pages 115-117). Further, it was intended to go beyond the peer-assessment "marking" part of the process, complementing this with its feedback or socially constructed learning component (Cushing *et al.*, 2011), through which skills are developed (Dochy *et al.*, 1999). Accordingly and despite students taking an active role marking the peer-assessment encounters, they were encouraged to use it as a framework for informed dialogic peer-feedback, thus focusing it more on the "learning" rather than on the "assessment" (Liu and Carless, 2006; Driessen and Scheele, 2013).

The dental and medical literature on peer-assessment methods contains many reports (Chapter 1) mainly focused on different aspects of validity, reliability and/or feasibility (Davis, 2002; Archer *et al.*, 2005; Dannefer *et al.*, 2005; Lurie *et al.*, 2006; Evans *et al.*,

2007; Archer *et al.*, 2008; Davies *et al.*, 2009; Mackillop *et al.*, 2011a; Speyer *et al.*, 2011; Taylor *et al.*, 2013; Ali *et al.*, 2014; Basehore *et al.*, 2014), and students' perceptions of their effects (Dannefer *et al.*, 2005; Shue *et al.*, 2005; Larsen and Jeppesen, 2008; Nofziger *et al.*, 2010; Setna *et al.*, 2010; Teich *et al.*, 2014). However, peer-assessment studies focusing on all five variables that define the utility of an assessment method according to van der Vleuten (1996), are lacking. To the authors' knowledge, this study is the first report to present the comprehensive analysis of a dental undergraduate peer-assessment and peer-feedback protocol's reliability, validity, feasibility, acceptability, and most usefully, its educational impact.

The absolute G coefficients for pre-clinical peer-DOPS (0.72) and that for clinical peer-DOPS (0.79) were higher than those obtained in the piloting of the peer-assessment protocol (0.62 and 0.67, respectively) (Chapter 3, page 124). However, the current Chapter G coefficients for pre-clinical peer-DOPS, clinical peer-DOPS and clinical peer-mCEX were lower than those previously reported (Moonen-van Loon *et al.*, 2013) for 16 encounters using DOPS (0.88), Mini-CEX (0.89), and MSF (0.88) for the assessment of 953 medical residents.

Notably, all our three formative peer-assessment instruments including pre-clinical peer-DOPS, clinical peer-DOPS and peer-mCEX reached the generally accepted reliable coefficients of ≥ 0.7 for low-stakes situations (Beard *et al.*, 2011) at 8, 8 and 7 encounters, respectively (Table 5-5), and thus the first hypothesis of this Chapter can be accepted: The implemented structured peer-assessment protocol is reliable.

This is in agreement with the conclusions of Williams *et al.* (2003) in that a minimum of 7 to 11 judgements of clinical performance are required for reliable findings. Coincidentally, our results also demonstrated the need for 10 or more peer-assessment and peer-feedback encounters to positively impact students' academic performance. This is probably the most important contribution of the present study: work-based, structured and continuous formative peer-assessment and peer-feedback of both pre-clinical and clinical dental undergraduate students has the potential to significantly enhance students' learning. In the words of Schuwirth (2013), the protocol does not only "make the horse drink" but also "make it want to drink".

This important finding might be explained by the requirement for students who engaged in the peer-assessment and peer-feedback protocol, to take an active role in the

management of their own learning (Liu and Carless, 2006). Thus, students self-regulate their learning by monitoring their work using their peers' feedback as a catalyst, and by expressing and articulating what they know or understand (Liu and Carless, 2006). This being consistent with social constructivist conceptions of learning in which social interactions with other learners play a fundamental role in the development of understanding and meaning, bridging the gap of the zone of proximal development through scaffolding (Wood *et al.*, 1976; Vygotsky, 1978) (further explained in Chapter 1, Constructivism in page 31). This is not to deny the possible implication of other learning theories like experiential learning (Kolb, 1984) or reflection in- and reflection on-action (Schön, 1983), explained in Chapter 1 (pages 33-35) and discussed below.

Following Vygotsky's model (1978), it can be hypothesised that the scaffolding peers provided to each other was, somehow, incremental. As shown in the graphical representation of Vygotsky's model presented in Figure 1-1 (page 32), the ZPD is the area where the learner can move with help of a more knowledgeable other: the scaffolding. In Chapter's 1 representation, the ZPD is wide, thinking of the strong scaffold a clinical tutor can provide to the learner, but the edge of the ZPD is a strong border, highlighting a clear limit where the student can go. Now, if we adapt this general model to the results of the present Chapter as interpreted in Figure 5-7, it can be noticed that the ZPD has been divided in smaller areas representing each peer-feedback encounter.

This, considering the "knowledgeable other" is a peer and her or his scaffolding cannot be compared to the tutor's one. However, the repetitive nature of the current peer-assessment protocol, allows the thinking of the possibility of the represented incremental scaffolding effect. Further, it can also be noticed in this adapted model the dotted edges of each incremental ZPD area in contrast to the strong border of the general model, representing the permeability of the development zone so that students 'can' go a bit further. Finally, the model has been drawn containing ten growing ZPD areas to represent the need for the same number of peer-encounters to obtain a measurable educational impact.

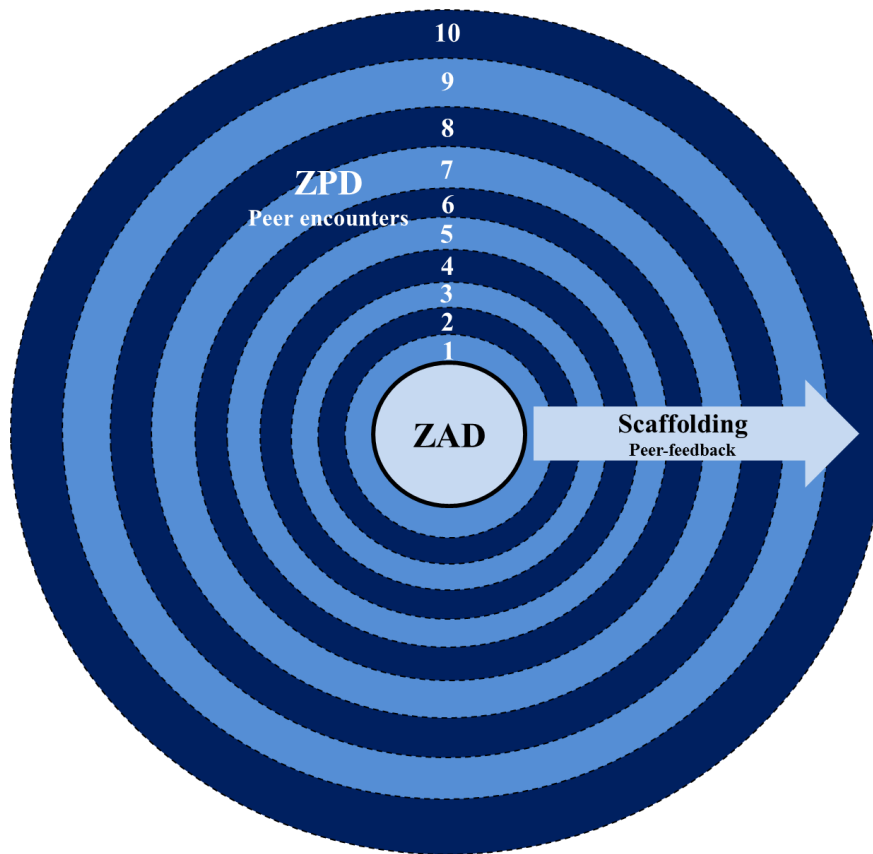


Figure 5-4 Graphical interpretation of Vygotsky ZAD (Zone of Actual Development) and ZPD (Zone of Proximal Development) (Vygotsky, 1978) according the results of the present chapter.

From a different perspective and based on Nicol and Macfarlane-Dick (2006) model of self-regulated learning, students in our study might have taken a proactive (rather than reactive) role in developing their objectivity in the continuous exercise of observing and providing feedback on their peers' work, which could have then been transferred to their own pre-clinical or clinical work, enhancing their learning (Falchikov, 1995), and potentially improving performance in high-stakes assessments (Liu and Carless, 2006).

This enhancement of student learning as a result of peer-assessment and face-to-face immediate dialogic peer-feedback on each other's work, has been reported to occur by means of diplomatic criticism, analysis and reflection (Falchikov, 1995; Dochy *et al.*, 1999).

The same finding may explain the statistically significant increase and positive effect size in reflective skills between those students who completed a higher number of peer-assessment and peer-feedback encounters as compared to those who did none, which again allows for the sixth hypothesis of the Chapter to be accepted: The peer-assessment

exercise stimulates students' reflective skills.

Consequently, keeping the focus on the learning effects of the protocol, the impact of the current peer-assessment and peer-feedback intervention was especially important for the former groups (those with ≥ 10 encounters). Accordingly, the group of 19 BDS 2 students who completed 10 or more peer encounters demonstrated both a significantly better performance and medium effect sizes in their high-stakes MCQs ($p=0.047$, $d=0.49$), Clinical Skills Examination viva ($p=0.050$, $d=0.48$) as well as practical ($p=0.21$, $d=0.56$) (Table 5-15), but also a statistically significant increase and large effect size in their Understanding ($p=0.011$, $d=0.71$) and Reflection ($p=0.0007$, $d=1.12$) skills (Table 5-19), as compared to those who did not participate in the peer-assessment exercise.

Indeed, MCQs were aimed at testing the degree of understanding of a concept or topic from different aspects of the course by knowledge recall, association and classification of concepts. Further, the Clinical Skills Examination practical judged the students' clinical competence in assessing and restoring a tooth in the phantom head and was thus quite similar to the continuous peer-assessment and peer-feedback protocol domains. The viva required students to answer clinical /materials /instrument questions related to a given clinical problem thus necessitating students to not only recall and understand, but also to apply knowledge in order to solve the clinical problem. With a different aim, written SAQs and SNQs are designed to assess knowledge recall which contrasts with the more application of knowledge focus of the pre-clinical peer-DOPS form. This might explain why the peer-assessment exercise did not produce any measurable effect in any of the studied groups. Further, going back to the preceding Chapter 4 (Table 4-7 in page 145), Paper 1a, consisting of short answer questions and short note questions for BDS 1 to 3, similarly did not correlate to any of the RQ constructs.

Interestingly, the group of 11 BDS 2 students who completed between 5 and 9 peer encounters embodied a large effect size of their Understanding ($d=0.77$) and Reflection skills ($d=0.52$) in a year time. However, this was not reflected in a better examination performance, perhaps suggesting the need for a longer period of continuous reflection.

In a similar trend, those 22 BDS 5 students who performed ≥ 10 peer-DOPS and/or peer-mCEX exercises exhibited a significantly better performance in their high-stakes clinical scenario Essays, MCQs and Clinical Reasoning Examinations, when compared

to the rest of their class. Furthermore, they also showed a significant increase in their Critical Reflection skills, which was not seen in any other group. This might indicate students reached the deep learning qualitative, “relational” and then “extended abstract”, phase of the SOLO taxonomy (Biggs and Tang, 2011 pp. 86-91), which has been described as a useful complement to the Kember *et al.* (2008) RQ when teaching and assessing understanding, construction and interpretation.

Once more, as observed in Chapter 4 (Table 4-7 in page 145), where the OSCE results correlated to both RQ Habitual Action and Understanding scales, this chapter's imperceptible impact of the peer-assessment protocol on OSCE's scores, was unforeseen. However, the answer to this might be found precisely in its positive relationship with Habitual Action and Understanding. As shown in Table 5-18, the peer-assessment and peer-feedback protocol only showed a significant difference and a large effect size ($p=0.003$, $d=1.04$) in the Critical Reflection scale for those BDS 5 students with 10 or more encounters. It would then be peculiar to find an increase in OSCE's scores when the evidence from the previous chapter showed they are only related to Habitual Action and Understanding. Furthermore, the same 22 students who peer-assessed each other 10 or more times and demonstrated a large Critical Reflection effect size, also showed a large negative size effect of $d= -0.65$ in their Habitual Action scores. This could be interpreted as a move from automated responses to becoming aware of why they perceive, think, feel or act as they do (Kember *et al.*, 2008).

BDS 2 Clinical Skills Examination viva section and the BDS 5 Essay and Clinical Reasoning Examinations have in common the requirement that students solve problems or cases. Thus, as presented by Mezirow (1991 pp. 99-117), this leads to higher order thinking, as the central dynamic involved in problem solving is reflection.

However, there still remains the question why BDS 2 students did not experience an improvement or higher effect size in their Critical Reflection skills, as did BDS 5 students. The answer to this might be found in the different environments both groups of students conducted the peer-assessment protocol, or rather what happened next. Following the trained protocol, and circumscribing it to Kolb's four-stages experiential learning cycle by which knowledge is created through the transformation of experience (Kolb, 1984 p. 41) (Chapter 1, Experiential learning in page 33), both groups of participating students observed/performed the pre-clinical or clinical task (Kolb's first stage: concrete experience), peer-assessed and debriefed the experience during the

course of the peer-feedback (Kolb's second stage: reflective observation), followed by an assimilation and distilling of the feedback information through self-assessment/reflection to draw new implications for actions (Kolb's third stage: abstract conceptualization). Finally, these new implications must be actively experienced for the learning to be effective (Kolb's fourth stage: active experimentation). However, in our protocol this stage of the cycle did not happen immediately after the abstract conceptualization, if it happened at all, as the third stage was performed at the end of each pre-clinical or clinical session. Hence, according to Kolb, conceptualized (3rd stage) but not experimented (4th stage) new implications/ changes/ frames may be missed (Kolb, 1984).

Returning to our protocol, BDS 5 students spent 68% of their time working with patients in the clinic (Chapter 2, Figure 2-4 in page 99), meaning their conceptualizations (3rd stage) could be experimented (4th stage) right after and in a daily basis. In sharp contrast, BDS 2 students' conceptualizations could only be experimented one week later, as they attended the pre-clinical skills laboratory on weekly bases. Further, according to Jarvis's (1987) revised Kolb's cycle, and due to this missing or delayed experimentation 4th stage of the latter students, it might be argued that they involuntarily adopted a "non-reflective learning" practice instead of the intended reflective learning. The author's model describes "non-reflective learning" as an acquisition of physical (manual) skills without reflection, which is the most common form of learning that occur in everyday life, but at the same time one that does not result in innovation (Jarvis, 1987). This contention might be supported by the present Chapter results in which participating BDS 2 students demonstrated the most significant results ($p=0.021$) and highest effect size ($d=0.56$) in their Clinical Skills Examination practical, which judges the students' practical skills to restore a tooth in the phantom head.

In a different but complementary theoretical perspective, BDS 5 participating students might have reached the critical reflection level of thinking, as they were working and practicing the peer-assessment protocol in the very relevant and problem-centred environment of the patients' clinic. Thus according to Dewey's theory of reflective thought and action (Dewey, 1909; Dewey, 1938), they might have been most motivated to reflect to resolve the conflict when facing the "inadequacy" or "state of uncertainty" (that is failures and difficulties) of clinical results (Dennick, 2012). At a higher level, if we think of this "inadequacy" as unusual or unexpected problems, surprises or even conflicts while treating a patient, they might have triggered the Schön (1983; 1987)

concepts of reflection-in- and reflection-on-action (Chapter 1, Reflective practice learning, in page 35), where participants examined the what, the how and the alternatives of the patient encounter. However, as he further states, in the same situation of reflection-in- and reflection-on-action of an ill-defined problem but adding the opportunity to share and process thoughts from the experienced problem with others (peers and facilitators), it is possible to reach the higher critical reflection level of thinking, questioning the why and the beliefs either helping or hindering the work at the encounter. Academically important, this is highly unlikely to happen without the pressure of inadequacy and facilitation through others (Miettinen, 2000; Stocker *et al.*, 2014).

There is no published data available to compare the educational impact of the current dental students' peer-assessment and peer-feedback protocol. However, a meta-analysis on 41 studies on medical clinical performance (Veloski *et al.*, 2006), reported that 32 of the reports demonstrated a positive impact of feedback, but the variation in outcome variables impeded any systematic analysis of effect sizes. Somehow further, the meta-analysis of feedback interventions by Kluger and DeNisi (1996) discussed in Chapter 1 (Feedback that works, in page 70) found the averaged effect size of feedback interventions across four groups of variables (cues, task characteristics, situational, and methodological) to be $d=0.38$, though peer-feedback or peer-tutoring was not within the moderators.

However, the huge study by Hattie (2012), made it possible to relate the present chapter's effect size results to his "list of influences on achievements". Thus, in his study the influence of feedback in classroom had an effect size of $d=0.75$ (especially when it is task focused and not to the self) and is ranked number 10 in a list of 150 influences, whereas "peer tutoring" has an effect size of $d=0.55$ and is tiered number 34. As it can be noticed, both interventions have an effect size well above the $d=0.40$ "gold standard" (Hattie and Timperley, 2007). Moreover, a just published meta-analysis of 31 studies on feedback for simulation-based procedural skills training which included six on dental students –though none on peer-feedback- found a positive effect size of $d=0.74$ in favour of feedback (Hatala *et al.*, 2014).

Accordingly, BDS 2 participating students who peer-assessed and provided peer-feedback to each other 10 or more times, exhibited an effect size of $d=0.80$ for the average of all their high-stakes examinations, as compared to those who did not

participate. Similarly, those BDS 5 participating students with 10 or more peer encounters revealed an effect size of $d=0.74$ for the average of all their Finals. Consequently, the fifth hypothesis of the Chapter is accepted: The implemented structured peer-assessment protocol has a positive educational impact.

Returning to the psychometrics of the peer-assessment protocol, notwithstanding that previous studies have reported the validity of the conventional formats of both Mini-CEX (Norcini *et al.*, 2003; Al Ansari *et al.*, 2013) and DOPS instruments (Wilkinson *et al.*, 2008; Delfino *et al.*, 2013), the new peer-DOPS and peer-mCEX rating scale and modified domains required a new validation process. Thus, content validity was ensured by the fact that all three forms were suitable for 99% of the encounters allowing a broad range of pre-clinical and clinical situations and procedures (Williams *et al.*, 2003) (Table 5-2 and Table 5-3).

Criterion validity, as a correlation between peer-assessment scores and high-stakes end-of-year examinations, was statistically significant for both peer-DOPS forms but not for the peer-mCEX, despite BDS 5 peer-DOPS and peer-mCEX scores were very similar. This inconsistency might be explained by the low number of 77 peer-mCEX completed encounters compared to 576 peer-DOPS encounters. It should be borne in mind that undergraduate dental education is based on strong procedural practice, initially in simulators (phantom heads) and subsequently on real patients (KCLDI teaching methodologies by course year in page 99). Clinical students perform a patient's clinical examination/assessment and implement a treatment plan during the successive clinical sessions. This explains the big difference between peer-DOPS and peer-mCEX numbers. From this perspective, it is probably correct to think that the inclusion of the peer-mCEX for undergraduate BDS 5 students was not well thought out.

Regarding construct validity, both groups of students using peer-DOPS and peer-mCEX forms demonstrated a consistent ability to detect significant differences in their peers' performance in the diversely assessed domains (Table 5-7, Table 5-8, Table 5-9), despite Sluijsmans *et al.* (2002) hypothesis that novices have a lower capability to assess certain domains. However, and as the same authors ascertained, training students in assessment skills positively affects their performance, which is exactly what our protocol did. Indeed, by using written/video examples and role-playing, all participating students in the present study received an intense training and familiarisation session on observation, peer-assessment, peer-feedback, action plan and completion of the peer-

DOPS and peer-mCEX instruments. As mentioned above, every peer-assessment session was monitored by the same researcher, who performed the training in a non-intrusive manner and thus any needed reinforcement was immediately done. Further, the educationally referenced scale containing a graphical and written anchor with the desired “increasing ability over time” (Chapter 3, Figure 3-6), might have had a positive impact in helping students to develop their assessment skills and explain their discrimination judgement. This possible indication of alignment between the scale and the students' expertise (Crossley *et al.*, 2011) might explain why participating students agreed or strongly agreed on the validity of the peer-assessment protocol in contrast with the study by Bennett *et al.* (2012) in which undergraduate medical students peer-assessed each other using the standard Mini-CEX norm-referenced scale (Below expectation, Borderline, Meets expectations and Above expectations).

Moreover, and possibly due to the same aforementioned reasons, students also revealed their capability to perceive significant changes in performance over time as early as the initial sessions. This was particularly interesting as the pattern was dissimilar between pre-clinical and clinical peer-DOPS and peer-mCEX scores (Figure 5-3). Thus, at the beginning of the academic year (September), BDS 5 peer-assessment scores were higher for patients' clinical examinations and treatment planning assessed using peer-mCEX than those given for clinical procedures by peer-DOPS. Later, both scores showed an upward gradient, though peer-DOPS' scores were steeper reaching those of peer-mCEX by November. This might indicate students felt initially more comfortable assessing patients than performing procedures. Further, by December/January both scores had reached a plateau (means 5 to 6) which continued as such until the end of the study. Surprisingly, these high and stable scores were reached during the same period as the students gave their Dental Foundation Training interviews which could be interpreted as they considered themselves ready as safe beginners.

In contrast, and likewise the same peer-assessment protocol preceding pilot study (Chapter 3, Figure 3-7 in page 126), BDS 2 students only started to notice peer improvement at the fifth session, while during the first four encounters scores dropped significantly. As mentioned before, this difference might be explained by an initial calibration process (Hauser and Bowen, 2009), the needed time to gain experience as evaluators (Karl *et al.*, 2011), and becoming at ease with making their work public (Liu and Carless, 2006), or even adjusting to the new pre-clinical learning environment (Schoenrock-Adema *et al.*, 2007).

Altogether, both undergraduate groups in the present study showed a similar capacity to perceive an increasing ability over time as that reported by Davies *et al.* (2009) in the already graduated foundation medical trainees' peer-assessment (using the mini-PAT), and that of Playford *et al.* (2013) and Prescott-Clements *et al.* (2011) in which staff trainers continuously assessed undergraduate medical students and postgraduate dental trainees, respectively, with similar clinical assessment forms. Consequently, and in light of the above discussed validity components of the protocol, the second hypothesis of the Chapter can be accepted: The implemented structured peer-assessment protocol is valid.

The overwhelming students' perception of the protocol's acceptability and fairness, expressed through the anonymous end-of-study questionnaire, was unexpected and allows to admit the fourth hypothesis of the Chapter: The implemented structured peer-assessment protocol is acceptable. In addition, no student stopped participating in the study or asked for the return of their completed peer-DOPS and peer-mCEX forms, having the freedom to do so as stated in their signed consent forms. As stated by Biggs and Tang (2011), this high acceptability could be interpreted as students "became accustomed to being observed by one another when they gave and received peer-feedback".

The explanation to this might be found in the other answers to the same anonymous questionnaire. Thus, students might have found the peer-assessment exercise acceptable as they largely scored it to be "easy and straightforward to use", and that it helped them "to identify learning needs and to improve their performance". Another possible reason can be found in the study by Shue *et al.* (2005), who concluded that "most students are willing to participate in peer-assessment as long as their preferences are taken into consideration". As a matter of fact, during the whole study period, students were always heard when they wanted to ask or give a comment on how to improve the protocol.

Besides the positive psychometric properties described above, there are also some practical reasons to implement a peer-assessment and peer-feedback protocol. Thus, in the presence of resource constraints and staff problems to provide sufficient feedback (Liu and Carless, 2006), peer-assessment and feedback can help to overcome the frequently reported lack of immediate feedback in workplace-based assessment as well as difficulties in finding a staff assessor (Quantrill and Tun, 2012). Furthermore, previous studies have reported a box ticking attitude when members of staff complete clinical assessment forms (Sabey and Harris, 2011), which was not observed or reported

in the present study.

Given these points and the present study's positive impact on students' academic results, supported by all other utility variables, peer-assessment and peer-feedback have significant potential in the current dental education setting. A frequently mentioned innovation in pre-clinical and clinical dental education is that of students working in pairs or teams mimicking the real working environment (Coomarasamy and Khan, 2004; Haden *et al.*, 2010) and assisting others to learn. Thus, the implementation of a peer-assessment and peer-feedback in such naturalistic settings would only mean structuring and organising a protocol without any changes in curricular content, which otherwise would mean losing an opportunity to foster students' academic and reflective skills. In other words, this would mean moving the serendipitous interactions between students, which would inevitably happen, from the uncontrolled "hidden" to the stated "formal" curriculum (Hafferty and Franks, 1994).

Future research may address some variables that this study did not contemplate. Peer-assessment and peer-feedback information could be used for an early exposure of those students needing tutor remediation. Further, as suggested by some students in the present study (Table 5-12), allocating specific time slots at the end of the session to allow peer-assessment and peer-feedback, and moving into electronic peer-DOPS and peer-mCEX forms, could attract more students or even allow them to complete more encounters. The latter student's suggestion might also help in easing the excessive amount of time and effort required to implement and monitor the current protocol. It would also be helpful to know whether peer-assessment or reflective skills scores have any predictive value in identifying the pre-clinical student who will do well in the clinic (Fugill, 2013) and in her/his future professional life.

5.7 Conclusions

The current study protocol of continuous, formative and structured pre-clinical and clinical peer-assessment as a framework for immediate dialogic peer-feedback of undergraduate dental students is not only reliable, valid, feasible and acceptable but it also has the potential to make a significant educational impact, provided it is practiced during ten or more peer encounters. In this instance, development of students' academic and reflective skills can be expected.

These results suggest that dental undergraduate students can be given more responsibility to take an active role in the low-stakes formative phase of learning by moving from being passive students to active trainers, as in the present peer-assessment and peer-feedback protocol, especially if our aim is to teach them to self-regulate and control their learning and so be prepared to manage their own education throughout life.

Thus, in this paradigm of enabling students' self-learning in parallel to a traditional delivered instruction, the next Chapter will further analyse the peer-feedback narratives students provided to each other to try to elucidate any clues that may help to better understand the present Chapter results and further develop the peer-assessment protocol.

Chapter 6

Analysis of Dental Students' Peer-feedback Narratives from a Prospective Peer-Assessment Protocol

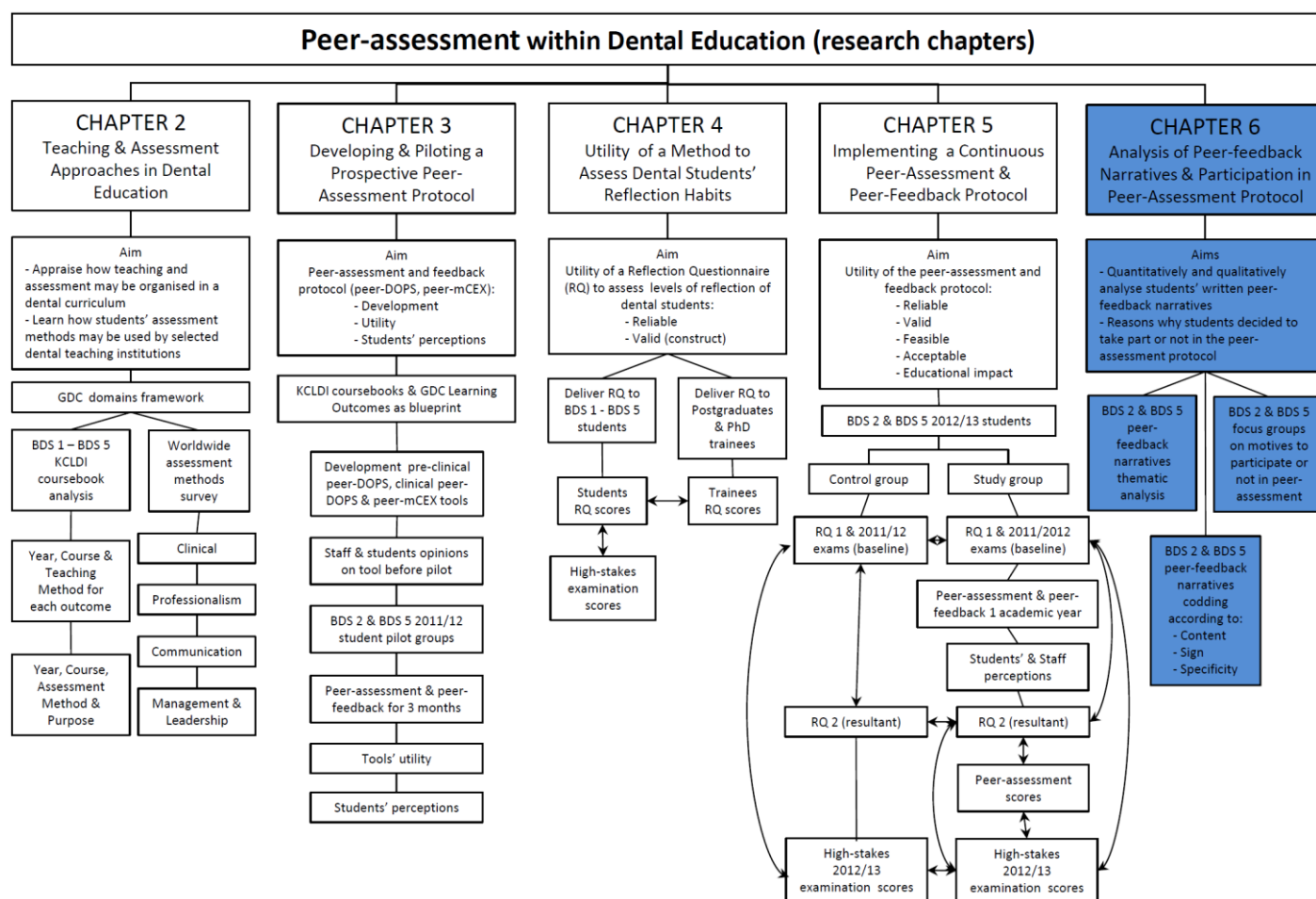


Figure 6-1 Flowchart of the study's research chapters. The sixth chapter is highlighted to help in providing an overall view.

6.1 Introduction

The move towards assessment *for* learning rather than assessment *of* learning has been described as “nothing short of a revolution in the conceptual framework of assessment” (Schuwirth and van der Vleuten, 2011b). In this approach assessment aims to drive the learning experience and so meaningful qualitative feedback is at the heart of the process (Assessment purposes in page 43) (Bok *et al.*, 2013).

However, there are many factors influencing the interpretation and uptake of such

feedback (Eva and Regehr, 2011). These factors include its positive or negative signs (Kluger and Van Dijk, 2010), students motivations (ten Cate *et al.*, 2004; Playford *et al.*, 2013), its credibility and nature of the task (Watling, 2014), evidence and specificity (Delva *et al.*, 2013), delivery moment and frequency (Hatala *et al.*, 2014), feasibility (Smither *et al.*, 2005), receptor focus (Kluger and Van Dijk, 2010), task perception (Pelgrim *et al.*, 2014) and emotional impact (Urquhart *et al.*, 2014).

The complex mixture of these and other factors can have a paradoxical effect (Eva and Regehr, 2011), reinforcing the notion that there is no simple recipe for the delivery of feedback (Boud and Molloy, 2013; McKenzie, 2013). Further, the interplay of aspects might explain that all feedback is not good feedback as it can both help or hinder learning (Kluger and DeNisi, 1996) (Chapter 1, Feedback in page 62).

Against this background and our findings in Chapter 5, we wanted to determine how our peer-assessment and face-to-face immediate peer-feedback protocol could have influenced students' skills. We therefore investigated the way students constructed, focused and based their feedback narratives.

6.2 Aim

The aim of this Chapter was to quantitatively and qualitatively analyse students' written peer-feedback narratives from the peer-assessment exercise, as well the reasons students decided to take part in, or not take part in the peer-assessment protocol.

6.3 Hypotheses

- i. Pre-clinical and clinical dental students participating in the peer-assessment protocol provide only general and positive peer-feedback in the clinical domain.
- ii. Participating students decided to take part in the peer-assessment protocol as a result of intrinsic and extrinsic motivations.
- iii. Students who decided not to take part in the peer-assessment protocol did so as a result of intrinsic and extrinsic motivations.

6.4 Materials and Methods

6.4.1 Quantitative coding of peer-feedback narratives

All 40 BDS 2 and 68 BDS 5 students (mean age=23.5, sd=2.6, 70 females, 38 males) who participated in the year-long, structured peer-assessment and peer-feedback protocol (Chapter 5), completed a total of 1169 peer-DOPS (N=516 pre-clinical and N=576 clinical) and peer-mCEX (N=77) forms. Of these, 1104 (94.4%) included a written feedback comment. These comments were transcribed to a word processor file, their words counted and the feedback narrative content coded by the same researcher (JT), according to the four GDC “Preparing for Practice” (2012a) domains (Clinical, Communication, Professionalism, Management & Leadership).

The peer-feedback narratives' sign was categorised as positive, when providing information about success; negative, when delivering information about failures; or mixed (Kluger and DeNisi, 1996). The narratives' specificity was also classified as task specific, when the procedure peers were performing was clear and/or gave a detailed comment on what and/or why something went well or otherwise; or general, when the procedure could not be identified and/or there was no reference to what and/or why something went well or badly. These factors were then analysed separately for BDS 2 and BDS 5 students using descriptive statistics.

6.4.2 Qualitative thematic analysis of peer-feedback narratives

To enable this analysis, the 1104 written feedback comment from BDS 2 and BDS 5 students were grouped into themes (by the same researcher), using an “open coding” interpretative process to break down the data analytically (Corbin and Strauss, 1990) and ensuring all manifestations of each theme had been accounted for (Pope and Mays, 2013). Feedback comments were then compared and those conceptually similar were labelled and grouped together to form themes. These emerging themes were categorised according to the four GDC domains (General Dental Council, 2012a).

6.4.3 Focus groups on the reasons peer-assessment participating and non-participating students decided to take part or not in the protocol

Of the 154 BDS 2 and 155 BDS 5 students invited to participate in the peer-assessment study, 40 (26% of the class) and 68 (44% of the class), respectively, agreed to take part in the protocol. For future implementation purposes, the reasons why students decided

to participate or not were investigated.

During April 2013 (before the peer-assessment and peer-feedback study ended), all BDS 2 and BDS 5 students were invited for a snack lunch, independently of their participation in the study, making it clear that the purpose was to gain in-depth understanding of the reasons that guided them to take part or not in the peer-assessment protocol. These 40-minute meetings were organised and facilitated by the same researcher involved in the study and were performed on two consecutive weeks on Tuesdays and Fridays for BDS 5 (completing 6 meetings) and Thursdays for BDS 2 (completing 2 meetings). Particular care was taken to ensure that each group contained students who participated in the peer-assessment protocol as well as those who did not. Meetings were not recorded or videotaped and the facilitator took written notes of the main themes appearing. Students signed a consent form which stated that their comments would be kept anonymous.

To stimulate the discussion, the facilitator opened each session by inquiring who of the students present did take part in the peer-assessment and peer-feedback protocol and who did not, and why?

As with the feedback narrative analysis, notes from the 8 focus group discussions were compared, labelled and grouped together in theme categories (Corbin and Strauss, 1990).

6.5 Results

6.5.1 Quantitative coding of peer-feedback narratives

Peer-feedback narrative mean number of words was 19.3 (range=2-44) for BDS 2 and 25.2 (range=6-56) for BDS 5. The most common BDS 2 content peer-feedback narratives related to the GDC Clinical domain (81%) and related to practical clinical skills (58%) and clinical knowledge (23%). In contrast, the most common BDS 5 content narrative related to the Management & Leadership domain (36%), closely followed by Communication (32%) (Table 6-1). It should also be noted that whilst most feedback comments related to a single domain, some did mention more than one domain competency, some examples of which are cited below.

Table 6-1 Proportional quantitative coding of BDS 2 and BDS 5 peer-feedback narratives according to their content, sign and specificity.

Peer-feedback Narrative Code		BDS 2	BDS 5
Content (GDC domain)	Clinical Practical	58%	12%
	Clinical Knowledge	23%	0%
	Communication	3%	32%
	Professionalism	1%	12%
	Management & Leadership	10%	36%
	Overarching competencies	5%	8%
Sign	Positive	91%	65%
	Negative	2%	13%
	Mixed	7%	22%
Specificity	Task specific	77%	74%
	General	23%	26%

BDS 2:

“Successfully placed on clamp, however need to remember to floss between each tooth and work a lot faster”

“Well time managed and very good manual dexterity and posture. Need to study more on tooth morphology & mixing materials”

“You really need to work on posture and go for harder challenges as it is the best way to learn. I also need it so let’s challenge each other every thursday”

“Practice more excavation as we just did to remove caries so you are sure in distinguishing between soft infected dentine and slightly harder affected dentine using an excavator”

“Continue practicing caries removal to enhance both time management and technique by improving assessment of depth of caries lesions as agreed on last session”

“Try and be quicker when carving and getting the right anatomical structure before the GIC sets”

“Very good use of time for the fissure sealants using all available materials, but should remember to use rubber dam”

BDS 5:

“Excellent work but need to be more confident”

“Don't be afraid to ask for help and remember to vaseline patient’s lips before placing disclosing tablet”

“Very well use of electrosurgery for the first time. However, don't forget time management”

“Difficult subgingival filling that entered into the pulp and you did well to notice this well in advance and asked for help. Remember to notice a pulp is necrotic from smell and do more tooth vitality tests before treatment”

“Try to get better in decision making and improve your confidence. We worked as a team but you will soon be alone”

“Remember to cover ultrasonic scales tip when not in use”

“Good communication with patient throughout the treatment and kept calm when struggling. You knew what you wanted to do, but as the treatment went difficult you should have asked for help at an early stage, don't compromise”

In relation to the ‘sign’ of peer-feedback narratives, 91% of those from BDS 2 students and 65% from BDS 5 students were categorised as positive, while 7% and 22%, respectively, being mixed. Some selected citations of the three sign categories (Table 6-1) are reproduced below:

Positive:

“Good foundation of understanding about the procedures. Suggest improving skills with hand pieces - positioning, time keeping and control of materials”

“You can do more than you think. Just be more confident in your own abilities”

Negative:

“At this stage you should already be able to work in a systematical manner”

“You lack management as should have booked double slot for this difficult RCT”

Mixed:

“Colour of the tooth nice, cusp pattern close but not morphologically correct”

“Excellent communication with patient, but if you wanted to do vitality test, make sure it is done before patient is anaesthetised. Think ahead!”

Similarly, most of BDS 2 and BDS 5 feedback comments were classified as task specific (77% and 74%, respectively) (Table 6-1). A few examples of these and other general comments were:

Task specific:

“You exhibited great competence using the incremental technique when restoring deep cavities. Continue working on and developing ability to finish composite restoration following the natural UL6 tooth morphology particularly in areas of cuspal build ups”

“After guidance was given from the tutor, you took them on board to extract the UR7 intact. However, you need to develop more clinical skills in XLA of a non-mobile tooth”

General:

“Improve knowledge of dental materials”

“Need to improve management”

6.5.2 Qualitative thematic analysis of peer-feedback narratives

All BDS 2 students' feedback narratives were grouped into 14 themes representing all GDC domains (General Dental Council, 2012a). There were 9 themes related to the Clinical domain, three to Management & Leadership, and one each to Communication and Professionalism (Table 6-2).

Table 6-2 Themes identified in BDS 2 peer-feedback narratives presented in alphabetical order (not frequency) and their corresponding GDC domain (General Dental Council, 2012a).

BDS 2 Themes	GDC domain
Caries removal	Clinical
Cavity preparation	Clinical
Communicate better	Communication
Correct procedure steps	Clinical
Dental material properties	Clinical
Efficiency	Management & Leadership
Harder challenges	Professionalism
Infection control	Clinical
Manual dexterity	Clinical
Practice more	Clinical
Restoration anatomy	Clinical
Time management	Management & Leadership
Working speed	Management & Leadership
Working position	Clinical

In contrast, BDS 5 students' feedback narratives were grouped into 24 themes representing all GDC (General Dental Council, 2012a) domains as well (Table 6-3). More specifically, sixteen themes were associated with the Clinical domain, six to Professionalism, two each to Communication and Management & Leadership, and one Overarching outcome.

Table 6-3 Themes identified in BDS 5 peer-feedback narratives presented in alphabetical order (not frequency) and their corresponding GDC domain (General Dental Council, 2012a).

BDS 5 Themes	GDC domain
Anticipate events	Professionalism
Ask for help	Clinical / Professionalism
Clinical skills	Clinical
Communicating with patient	Communication
Decision making	Clinical
Diagnostic skills	Clinical
Compromising patient	Professionalism
Efficiency	Management & Leadership
Informed consent	Clinical / Communication
Infection Control	Clinical
Knowing your limits	Clinical / Professionalism
Laboratory work evaluation	Clinical
Local anaesthesia	Clinical
Managing complications	Clinical
Managing patient's expectations	Clinical
Overall view	Overarching outcome
Organisation	Clinical
Patient anxiety	Clinical
Professional attitude	Professionalism
Self confidence	Clinical / Professionalism
Time management	Management & Leadership
Understanding procedure	Clinical
Vitality test	Clinical
Working speed	Clinical

6.5.3 Focus groups on the reasons peer-assessment participating and non-participating students' decided to take part or not in the protocol

A total number of 36 (9 BDS 2 and 27 BDS 5) students attended the focus group meetings. Of these, 25 (6 BDS 2 and 19 BDS 5) followed the peer-assessment protocol and 11 (3 BDS 2 and 8 BDS 5) did not participate. The atmosphere in the meetings was

always very positive and participants were honestly willing to give their views.

Those 25 students who participated in the peer-assessment and peer-feedback protocol gave a wide range of reasons to engage (Table 6-4). From these, there were three themes which emerged in all 8 meetings: “enjoy teaching”, “the need for more feedback”, and “trust my peers’ assessment”.

Table 6-4 Themes identified during the focus groups as the reasons why students who followed the peer-assessment protocol decide to take part in it. Recurrent themes are shown in bold.

Themes from participating students
Enjoy teaching
Induction tutorial was convincing
Informal peer-tutoring was done anyway
Study seemed innovative
The Kindle Fire HD incentive
The need for more feedback
The research certificate for portfolio
To help the PhD student in charge
Trust my peers’ assessment
Wanted to improve and develop clinical performance

By contrast, the 11 students who did not participate in the peer-assessment study but attended the snack lunch focus groups provided a limited number of reasons why they had not participated in the protocol. However, a common theme was in relation to bad prior formal experience of work-based assessment and feedback (though not necessarily coming from peers), particularly from those students with a previous university degree (Table 6-5).

Table 6-5 Themes identified during the focus groups as the reasons why students decided not to take part in the peer-assessment study. Recurrent theme is shown in bold.

Themes from not participating students
Bad previous experience with WPBA and/or feedback
Didn’t think it was helpful
Didn’t want to take the responsibility
Felt unable to assess peers
Have an iPad

6.6 Discussion

This Chapter reports the quantitative and qualitative analysis of students' written peer-feedback narratives from all those who exercised the peer-assessment protocol. It has also provided some insight into the reasons students decided to take part in, or not take part in the study.

The quantitative coding of both BDS 2 and BDS 5 peer-feedback narratives included their content, sign and specificity. The first analysis (content) of the results showed a notable difference between pre-clinical and clinical students' narratives. In a way, this might be expected as the environments where each group of students performed their work is totally different: BDS 2 in the pre-clinical skills laboratory with its manual dexterity challenges, and BDS 5 in "the real" world clinic with time and patient stresses. However, this may not fully explain the difference. The overwhelming 81% of Clinical domain in BDS 2 narratives, contrasts with the much lower 12% in BDS 5 ones and could be explained by the level of training of both groups of students. Hence, pre-clinical students are at the beginning of the clinical skills ladder whilst senior students are much further up. Additionally, the fact that BDS 5 students had to work with a patient clearly influenced the content of the provided feedback, focusing it more on other needed skills like Communication, Management & Leadership and Professionalism. This students' ability to provide peer-feedback not only in the clinical domain already allows for the first hypothesis to be rejected: Pre-clinical and clinical dental students participating in the peer-assessment protocol provide only general and positive peer-feedback in the clinical domain.

Unfortunately, there is a notable absence of evidence on the "content" of feedback studies (Hatala *et al.*, 2014) with which to compare our results. However, the evidence from the "Guidance Hypothesis" from the motor learning literature (Schmidt and Lee, 2011 as cited by Hatala *et al.* 2014), shows that terminal feedback - given at the end of the practice - is superior to concurrent feedback - given during the practice - for long term skill retention, as the latter may lead to an over-reliance on feedback (Hatala *et al.*, 2014). These findings support our peer-assessment and peer-feedback protocol as the feedback was provided at the end of the pre-clinical or clinical session. This was encouraged by our protocol requiring participating students to peer-assess the peer trainee following a structured framework which was meant to inform the subsequent feedback. Under these circumstances, they were unconsciously guided to provide

terminal feedback. Notwithstanding, and after reading the sample feedback narratives given above, it would have been inappropriate to do so in front of the patients and as instructed, students provided feedback privately. However, this terminal feedback is not to be confused with the constructivist scaffolding provided *during* the procedure (Wood *et al.*, 1976; Vygotsky, 1978).

The second quantitative peer-feedback narrative analysis coded them according to their sign as positive, negative or mixed. Once again, BDS 2 and BDS 5 students showed notable differences. Thus, whilst the vast majority of both pre-clinical and clinical peer-feedback narratives were coded as positive (91% for BDS 2 and 65% for BDS 5), BDS 5 negative ones were more than six fold those from BDS 2 (13% versus 2%). Similarly, mixed sign narratives from clinical students were three fold those from pre-clinical ones (22% versus 7%). These differences might be attributed to the contrasting pre-clinical and clinical contexts in which peer-feedback was provided. “Real-life” clinical training at dental schools involve being responsible for a patient, and so are very intensive and demanding and students usually feel stressed and tired (Kossioni *et al.*, 2012; Ostapczuk *et al.*, 2012; Tomás *et al.*, 2013).

However, the higher proportions of negative and mixed narratives signs from clinical students should not instinctively be considered damaging or unhelpful, especially if their attention was focused to the “task” being assessed and not as a threat to the self-esteem (Kluger and DeNisi, 1996). Then, a possible explanation for a positive effect of a negative signed peer-feedback might be found in the “Regulatory Focus Theory” (Higgins, 1997), which extends the hedonistic principle that humans are motivated to approach pleasure and avoid pain. Thus, Tory Higgins distinguishes two systems of self-regulation. On the one hand, a “promotion focus” concerned with accomplishments and aspirations which is motivated by achievement of rewards (things we *want* to do), and on the other, a “prevention focus” concerned with safety and responsibilities that is driven by the avoidance of pain or punishment (things we *have* to do) (Higgins, 1997; Watling *et al.*, 2012b).

Subsequently, using this Regulatory Focus Theory, Kluger and Van Dijk (2010), explained why “neither positive nor negative signed feedback affects performance in a constant manner”. Consequently, they showed that a positive signed feedback contributes to motivation and performance under a “promotion focus”, for example in tasks requiring creativity. By contrast, a negative signed feedback is encouraging under

a “prevention focus”, for example in tasks requiring vigilance and attention to detail, but this is debilitating under “promotion focus”.

Furthermore, Van Dijk and Kluger (2011) reported a large effect size ($d=0.67$) in performance of positive feedback among people working on promotion tasks (e.g. initiating changes, planning a task, challenging decision making, creative problem solving, assimilating new technology, presenting various alternatives, and generating ideas), but they also described a hampering effect size ($d= -0.37$) in performance when a positive sign feedback was given to individuals working on prevention tasks (e.g. detecting errors, maintaining safety, quality control, supervising, and work scheduling). Coincidentally, the former effect size for positive feedback in promotion focus tasks ($d=0.67$) resembles the ones from those BDS 2 and BDS 5 students who completed 10 or more peer-assessment encounters got in their average of all examination ($d=0.80$ and $d=0.74$, respectively) described in Chapter 5 (Table 5-15 in page 180 for BDS 2 and Table 5-16 in page 181 for BDS 5). Moreover, the current Chapter's BDS 2 facts in that 91% of their feedback narratives were coded as positive, 77% were task specific, and 81% with a clinical content, amalgamated together with Van Dijk and Kluger (2011) descriptions of promotion tasks, it is conceivable that these BDS 2 students worked mostly in a “promotion focus”. This because they were generally involved in tasks such as “initiating changes” from theory to practice, “planning a task” to successfully accomplish it during the pre-clinical session, and “assimilating new technology” like drilling handpieces, ultrasound and x-ray machines. If we further consider the results from Chapter 5 where the same BDS 2 students' perceived their pre-clinical tasks complexity mostly as “moderate” (66.7%) and “high” (14.1%) (Table 5-2 in page 165), it is again imaginable that the meaning of the tasks being practiced would incline them to adopt a “promotion focus”.

Regarding BDS 5 students, the coding of their peer-feedback narratives (content, sign and specificity) as well as their perceived clinical tasks complexity suggests the adoption of mixed “promotion” and “prevention” foci, as mentioned by Kluger and Van Dijk (2010) for medical staff. Firstly, considering that the majority of their feedback narratives were coded as positive and task specific, diversified in content domain (Table 6-1), and that these students' mostly perceived clinical tasks complexity as “moderate” (42.6%) and “high” (25.4%) (Chapter 5, Table 5-3 in page 166), it insinuates these senior students adopted a “promotion focus”. Furthermore, and according to Van Dijk and Kluger (2011) promotion tasks list, they were making “challenging decision” with

professionalism, embracing a “problem solving” attitude for *clinical* struggles, “presenting various alternatives” while *communicating* with patients and trainers, “generating ideas” while writing a treatment plan, and practicing *management* when “planning tasks”.

Secondly, however, these BDS 5 students were also performing tasks “they *have* to do” adopting a “prevention focus” (Higgins, 1997). Accordingly, while treating their patients, and especially when performing those dental crowns, dentures and root canal treatments perceived as highly complex (Chapter 5, Table 5-3 in page 166), they must have focused on detecting *clinical* errors, maintaining their patients’ safety, *managing* the quality of the procedures and *professionally* supervising each other, described as prevention tasks (Van Dijk and Kluger, 2011). Expanding this further, a negative signed peer-feedback (as the 13% in BDS 5 students) for this tasks requiring vigilance and attention to details, might have been encouraging for the students working on a “prevention focus” as described by Kluger and Van Dijk (2010). Further, under this kind of circumstance, Watling *et al.* (2012b) suggested that negative feedback (criticising or correcting a behaviour) is not only acceptable but desirable and more valued than a positive one (affirming or reinforcing a behaviour).

The above discussed assumptions of the present study results are consistent with Kluger and Van Dijk (2010) judgment in that “medical staff are faced with a mix of prevention and promotion foci”, especially in the case of our clinical BDS 5 students. However, our evidence is in complete disagreement with the same researchers speculation that the medical “community seems to emphasise error avoidance, risk management and the minimising of losses over creativity and other promotion-focused goals”. Consequently, it is not credible to base these statements just on a PubMed search comparing references with “diagnostic error” and “error reduction” (90.000) to those with ‘creativity’ and ‘opportunity’ (60.000) (Kluger and Van Dijk, 2010). In contrast, our data is in agreement with the detailed qualitative study by Watling *et al.* (2012b) in which they found that 22 medical learners were not “inclined to emphasise one focus over the other”. Further, the same task often included both foci and played a significant role in the workplace.

The qualitative thematic analysis of BDS 5 students’ peer-feedback narratives was, as expected, more clinically focused and patient-centred than those from BDS 2 students. Both groups of students provided peer-feedback to each other in a wide range of

Clinical, Communication, Management and Professionalism themes relevant to their particular pre-clinical and clinical working environments. However, four themes emerged at both groups' narrative analysis. These were efficiency, infection control, time management and working speed, signalling these to be ongoing learning skills in spite of the difference between both pre-clinical and clinical courses.

In a mixed qualitative - quantitative methods study by Nofziger *et al.* (2010), second and fourth year medical students provided narratives about their prior peer-feedback experiences where both groups expressed the same themes. However, second-year students reported higher interpersonal style feedback than fourth-year ones (71% versus 63%), whilst conflicting feedback was reported higher in senior students (7% versus 1%). Among similar reported themes from our subjects, these medical students remembered having received feedback on "being too quiet", "harder challenges", "overconfidence", "knowing your limits", "low self-esteem" and "low self-confidence". Another study on third-year medical students this time by White and Sharma (2012), reported similar written peer-feedback themes as our clinical BDS 5 students. Among these, they mentioned information management, initiative & self-direction, asking questions to learn, confidence, compassion, and respect for others.

The third part of the present Chapter's results where focus groups were conducted, attempted to clarify why only 26% of BDS 2 and 44% of BDS 5 students decided to take part in the peer-assessment and peer-feedback exercise, but at the same time provide some insight into why, for future improvements, the majority of BDS 2 (74%) and BDS 5 (56%) students decide not to do so.

Thus, among the first group of students - those who peer-assessed each other – some expressed their desire to improve and develop their clinical performance practicing the peer-assessment protocol. As described by Lurie *et al.* (2006), this could be a sign of senior students' mature judgements, as well as a positive ongoing process of students' developing their professionalism, described by Roff *et al.* (2011) as "proto-professionalism".

Another theme expressed by students as a reason to participate in the protocol that deserves some discussion is "the need for more feedback". The motives for this might be found in the results of the 2013 UK NSS. Thus and despite that a high 80% of the KCLDI "BDS Dentistry" and 88% of "BDS Dentistry Graduate/Professional Entry

Programme” surveyed students (UK average 70%), agreed with the statement “Feedback on my work has been prompt”, a lower 61% and 63%, respectively (UK average 64%), agreed that “Feedback on my work has helped me clarify things I did not understand” (Unistats, 2014). This could be interpreted as whilst students are being provided with prompt staff feedback it is not being perceived by some students as helpful. This has been described in the literature as an effect of feedback provided to passive recipients (Teunissen *et al.*, 2009).

Hence, the option of participating in the peer-assessment protocol where peer-feedback was conceived more as an accessible communication, interaction or dialogue between peers (Nicol and Macfarlane-Dick, 2006; Carless, 2007; Nicol, 2010), could have been seen as an opportunity to fill this gap.

Peers' trust was another frequently mentioned theme which helps to understand the high proportion of participating students that found the protocol acceptable and fair (85% of BDS 2 and BDS 5) as described in the previous Chapter (Table 5-10 in page 170). Similarly, Mann *et al.* (2011) described peer-feedback that took place in a relationship of mutual trust and respect, was definitely enhanced. Further, this positive atmosphere also contributes to explain the positive influence peer-feedback had in participating students' reflective skills and examination performance. Thus, as concluded by Watling *et al.* (2012a), credibility is a key factor for feedback to be influential. As they stated, only feedback that survives a critical assessment of the source credibility is likely to impact students' learning. If not credible, feedback will probably fall on deaf ears (Bing-You *et al.*, 1997).

Students enjoyed teaching each other. This is probably related to the increased students' motivation as a product of the cooperative learning approach which has been described as ideal for courses that require students to learn skills that require manual dexterity, knowledge and clinical reasoning (Asghar, 2009). Further, a positive peer-assessment and peer-feedback experience, despite its challenges (Walker, 2014), might have improved students' belief in themselves, increasing their confidence and self-efficacy (Bandura, 1997). Moreover, students' cooperation during the peer-assessment protocol might have benefitted the feedback receiver stimulating her or his reflection as discussed in Chapter 5, but also the provider as a peer-teacher. Hence, according to Dales's Learning Pyramid (Dale, 1946), listening leads to 5% recall, whereas teaching others, leads to 80% recall (ten Cate and Durning, 2007).

Additionally, this enjoyable cooperative experience based on peer-assessment and peer-feedback that was immediate, constructive, dialogic and focused to improve the performed task, might be a good example of socially constructed learning reflecting Vygotskian schools of thought (Vygotsky, 1978; Topping, 2005). This interaction between peers in an amenable way (Asghar, 2009) would be responsible for the benefits on the cognitive development produced in both assessor and assessed students, over learning in isolation (Vygotsky, 1978).

Despite the above mentioned intrinsic motivation themes, there also were some students that honestly stated some extrinsic reasons to participate in the peer-assessment study as they were motivated by the Kindle Fire HD prize draw, the Research Certificate for their portfolios, and even some others that wanted to help the PhD student conducting the study. Besides helping to augment the study sample, these extrinsic reasons together with the above intrinsic ones, allow us to accept the second hypothesis of this Chapter: Participating students decided to take part in the peer-assessment protocol as a result of intrinsic and extrinsic motivations.

Those students who chose not to peer-assess each other – but attended the focus groups – presented a few reasons, mostly intrinsic but extrinsic as well, for not taking part in the peer-assessment protocol. This, besides allowing us to accept the third hypothesis of the Chapter in that “students who decided not to take part in the peer-assessment protocol did so as a result of intrinsic and extrinsic motivations”, gave us a helpful insight for future implementations of the protocol.

Consequently, non-participating students talked about their uncomfortable earlier experiences with feedback while following their previous university degrees. This was a surprising revelation but was concurrent with the large number of students who already held a university degree and were now following the BDS programme. Thus, a total 26 BDS 2 (17% of the class) students and 37 BDS 5 (24% of the class) were part of the Dentistry Graduate/Professional Entry Programme or Dentistry Entry Programme for Medical Graduates. In agreement with this, a qualitative study conducted by Beaumont *et al.* (2011), investigated the impact of prior experiences of peer-assessment in higher education, and found a great proportion of students with negative experiences relating to trust, competency and plagiarism. Further, in the clinical environment Quantrill and Tun (2012) reported more than half of the 41 surveyed foundation medical trainees were not satisfied with the quality of the received feedback when assessed by staff with three

different forms of workplace-based assessment. In line with this and even more dramatic, Sabey and Harris (2011) surveyed results of 52 specialist medical trainees using workplace-based assessment showed that though they valued feedback, this was lost as it often came too late, it was of poor quality, and doubtfully honest, particularly when given face to face. Further, just now Urquhart *et al.* (2014) reported fifth year undergraduate medical students' feedback experiences as a process that happened 'to' them rather than 'with' them, one-way process, and metaphorically described "feedback as war". In light of this, the reason stated for choosing not to participate in the peer-assessment protocol based on bad previous experiences appears reasonable.

Another theme identified by those students who did not participate in the peer-assessment protocol, related to their reluctance to take the responsibility of assessing their peers. This feeling was also described by Brown and Glasner (1999), who reported the "common belief that assessment is the teacher's responsibility and some students resent being required to do the teacher's dirty work". Others felt unable to assess their peers which could be related to the Biggs and Tang (2011) statement in that peer-assessment can be stressful to some students. Both these themes probably denote a failure in the protocol's explanation and training to the students, as the "learning" of the peer-assessment process should have been emphasised over the "assessment".

However, this specific issue must not make us lose the overall view of the problems confronted by the current approach to feedback in higher education. Our results are further evidence supporting a need for a change in the quality (Sargeant *et al.*, 2008; Pelgrim *et al.*, 2012b), quantity (Nicol and Macfarlane-Dick, 2006; Quantrill and Tun, 2012; Urquhart *et al.*, 2014) and general approach (Walsh *et al.*, 2009; Archer, 2010; Hatala *et al.*, 2014) to improve the provision of a much needed feedback. In this perspective, this Chapter results have shown that peers can be of great assistance in providing valuable feedback to change students' behaviour.

6.7 Conclusion

This Chapter's analysis of peer-feedback narratives has shown that both pre-clinical as well as clinical dental undergraduate students, using their respective peer-DOPS and peer-mCEX forms as frameworks, were able to engage in a dialogic feedback focused to different domains, and of contrasting signs and specificities. However, while pre-clinical students' feedbacks were mostly focused on the Clinical domain, positive and

task specific, clinical ones were more concerned with Management & Leadership and Communication skills, and at the same time dared to provide more negative or mixed feedback keeping it task specific.

Thus, as emerged narrative themes resembled those differentiated items from their respective pre-clinical and clinical peer-assessment forms, it appears these provided a framework to help students in the provision of different feedback codes according to their specific work based circumstances.

Training on peer-assessment and peer-feedback should emphasise a collaborative and formative approach to learning in order to integrate those students with previous negative experiences with feedback. Further, the protocol should include intrinsic as well as extrinsic motivations in order to increase students' participation.

General Discussion

The educational literature supports peer-assessment and particularly its feedback component as a way of fostering student learning. The development, piloting and implementation of the current peer-assessment protocol demonstrated positive results which are summarised below (Table GD-1). However, despite its feasibility, peer-assessment is generally lacking in undergraduate education (Boud, 1991).

Based on Barr's adaptation (2000) of the four levels of educational intervention's hierarchy (Levels 1, 2a, 2b, 3, 4a and 4b) according to Kirkpatrick (1967), the academic effects of the studied peer-assessment protocol in students who completed 10 or more peer encounters can be coded as Level 2b: Acquisition of knowledge/skills. The hierarchy relates this to the "...acquisition of concepts, procedures and principles of inter-professional collaboration", and/or the "...attainment of thinking/problem-solving, psychomotor and social skills linked to collaboration" (Barr *et al.*, 2000 p. 10).

Similarly, the development of students' reflective skills allows this part of the intervention to be categorised as Level 3: Change in behaviour, which implies a "behavioural change transferred from the learning environment to the workplace prompted by modifications in attitudes or perceptions, or the application of newly acquired knowledge/skills in practice" (Barr *et al.*, 2000 p. 11).

These findings are important as most studies in medicine and dentistry provide educational outcomes limited to Level 1 (learner's reactions) and Level 2 (2a - modification of attitudes and perceptions, and 2b -acquisition of knowledge/skills), and only occasionally Level 3 (Change in behaviour) (Gordon and Findley, 2011; Hendricson, 2012). In addition, a review on learner outcomes in healthcare interprofessional continuing education (Gillan *et al.*, 2011) found 22% of the outcomes at Level 1; 66% at Level 2a; 1.4% at Level 2b; 10.2% at Level 3; and only 0.4% at Level 4 (4a -change in organisational practice, and 4b -benefits to patients).

Pedagogically, the most important contribution of the present thesis is the evidence that a carefully and rigorously planned longitudinal peer-assessment and peer-feedback protocol of dental undergraduate students can significantly enhance their academic achievements and develop their reflective skills (Chapter 5). Using Biggs and Tang adagio (2011 p. 51), the former allows them to fish for today's meal; the latter is the net that provides meals for the rest of their lives.

Table GD-1 Aims and outcome summary for each research chapter of the thesis.

	Aims	Outcome summary	Main implications
Chapter 2	<ul style="list-style-type: none"> To investigate the different teaching and assessment methodologies used in dental education using KCLDI as a true model. To investigate assessment practices and strategies used by selected dental teaching institutions around the world to measure students' progress particularly in relation to the use of peer-assessment. 	<ul style="list-style-type: none"> KCLDI comprises 1716 teaching sessions with the majority being lectures (40%) and practical & clinical (38%). Assessment includes 23 different methodologies and 153 sessions (55% formative, 30% summative, 15% hurdle). OSCE was the most commonly used assessment methodology (53% of institutions), followed by MCQ (40%), preclinical observation (32%), and WPBA (32%). Peer-assessment was used by 11% of the institutions. 	<ul style="list-style-type: none"> Similar assessment strategies are shared by dental teaching institutions from different countries and continents. New clinical assessment techniques are being employed in undergraduate dental education. The proportion of methods that assess critical thinking and problem solving is small.
Chapter 3	<ul style="list-style-type: none"> To develop and pilot a structured protocol of formative, prospective peer-assessment as an informed framework for immediate peer-feedback. 	<ul style="list-style-type: none"> The developed peer-assessment protocol is reliable, feasible, and well accepted by students. 	<ul style="list-style-type: none"> It is worth implementing the protocol in a larger sample followed for a longer period to provide generalizable data.
Chapter 4	<ul style="list-style-type: none"> To assess the reliability and construct validity of a self-reported reflection questionnaire as a method of assessing dental students' and postgraduate trainees' reflection habits. 	<ul style="list-style-type: none"> Kember's quantitative Reflection Questionnaire is a valid and reliable instrument to assess dental students' reflective skills. The higher the academic ladder the more reflective students are. Those ≥ 24 years of age and with a previous university degree demonstrated higher reflective habits. There is no gender difference in reflective skills. 	<ul style="list-style-type: none"> The diagnostic reflection tool can be used as pre- and post-intervention of the peer-assessment and peer-feedback protocol.
Chapter 5	<ul style="list-style-type: none"> To implement and assess the reliability, validity, feasibility, acceptability and educational impact of the peer-assessment and peer-feedback protocol. 	<ul style="list-style-type: none"> The protocol is reliable, valid, feasible and acceptable. It has the potential to make a significant educational impact, provided it is practiced during ten or more peer encounters. 	<ul style="list-style-type: none"> Dental students' academic and reflective skills can be developed through a structured and longitudinal peer-assessment and peer-feedback protocol, provided it is exercised at least ten times during the academic year.
Chapter 6	<ul style="list-style-type: none"> To quantitatively and qualitatively analyse students' written peer-feedback narratives. To scrutinise the reasons why students decided to take part in, or not take part in the peer-assessment protocol. 	<ul style="list-style-type: none"> Pre-clinical students' feedbacks were mostly Clinical, positive and task specific Clinical ones were more concerned with Management & Leadership and Communication skills, and provided more negative or mixed feedback keeping it task specific. Bad previous peer-assessment experiences precluded students to participate in the peer-assessment protocol. 	<ul style="list-style-type: none"> Peer-assessment framework help students to provide feedback focused to different domains, and of contrasting signs and specificities. Students' training on peer-assessment and peer-feedback must stress its collaborative and formative approach to integrate students with previous negative experiences. Intrinsic and extrinsic motivations to participate should be included.

The enhancement of students' skills is the result of an intervention that deliberately intended to increase the scope, frequency and, if possible, the depth of participants' feedback, to complement feedback tutors can reasonably provide (Evans, 2013). The powerful effects of such peer-feedback protocol encourage the organisation of collaborative student working (Ladyshevsky, 2013).

Taken as a whole, the results of the peer-assessment and peer-feedback protocol meets the criteria that effective learning needs to be *engaging*, *contextualised*, *active*, and *social* (Chickering and Gamson, 1999; Forsyth, 2003; Kaufman, 2003; Biggs and Tang, 2011; Harden and Laidlaw, 2013).

The peer-assessment protocol achieved this by requiring students to *engage* in two different ways:

In a constructivist model they connected and engaged the actual concrete experience of simulation (pre-clinical BDS 2 students) and clinical performance (clinical BDS 5 students) to previous similar events to activate what they already knew, had experienced and felt (Dennick, 2012; Pritchard, 2013).

It resulted in "observing students" engaging with the task of their peers. For example, the observing students classified the complexity of each procedure (Chapter 5, Table 5-2 in page 165 and Table 5-3 in page 166), accurately marked their peers performance (Chapter 5, Table 5-11 in page 171), and provided a "task specific" feedback in 77% of BDS 2 and 74% of BDS 5 encounters (Chapter 6, Table 6-1 in page 205). For them to set the case complexity, reliably assess their peers, and provide detailed feedback narratives they had to engage in their peers' tasks. It is also arguable that all students - participating and non-participating - were required to engage in their own respective tasks.

A possible explanation for this positive engagement effect might be the alignment (Biggs and Tang, 2011) of the peer-assessment domains, from both peer-DOPS and peer-mCEX, with the students' respective course learning outcomes and therefore their formal assessment (Nofziger *et al.* (2010), as shown in Chapter 3 (Table 3-1, pages 115-117). In contrast, a lack of alignment might have led to not engaging with the peer-feedback protocol (Ladyshevsky, 2010).

Similarly, the need for learning to be *contextualised* was addressed as both groups of participating students worked situated, be it in a pre-clinical simulation or in clinical

practice, in what are the day-to-day dental procedures. Whilst non-participating students were also doing this, peer-feedback narrative analysis from Chapter 6 demonstrates how the assessment domains of each of the BDS 2 and BDS 5 forms described in Chapter 3, might have helped both observing and performing students to contextualise the task itself not only as the required pre-clinical or clinical procedure but as a broader learning opportunity.

Consequently, it is hard to see if it was not for the peer-DOPS or peer-mCEX frameworks, how BDS 2 students put the very pre-clinical task in context, for example with working speed, time management and communication (Chapter 6, Table 6-2 in page 207). Further, BDS 5 students spoke about knowing their limits, efficiency, professional attitude and anticipating events (Chapter 6, Table 6-3 in page 208). This supports the idea that the peer-assessment experience demanded the students to be *active* contributors to their learning process as none in the working pairs could be simple passive spectators (Chickering and Gamson, 1999). It is well known that active learning results in a deeper processing of the learning experience (Divaris *et al.*, 2008; Harden and Laidlaw, 2013; Nicol *et al.*, 2014).

Moreover, by actively observing, scaffolding, assessing and providing feedback to each other in a social constructivist model (stressing social interaction), they had to ‘share’ what they know, what they don’t know, relate it to previous pre-clinical or clinical experiences and apply it to the current simulation or real patient treatment. As a result, they learnt by awaking developmental processes that only operate when interacting with their peers (Vygotsky, 1978). Further, the feedback was timely and immediate which, as discussed in Chapter 1 (page 71), is an important factor for effective feedback (Subramanian *et al.*, 2013), particularly for novices who need to learn things quickly, in context, and when still engaged with the task (Ladyshevsky, 2013). Students who were not involved in the study similarly had to actively work in their tasks but the difference was the relative absence of ‘share talk’ (Vygotsky, 1978) or dialogic feedback (Beaumont *et al.*, 2011; Pritchard, 2013) participating pre-clinical and clinical partners performed during and after their tasks.

In addition, whilst all students (both BDS 2 and BDS 5) were “actively engaged” in their own pre-clinical or clinical tasks, those who participated in the peer-assessment exercise benefited from the effects of being both observer (assessor) and trainee. As discussed in Chapter 1, previous research has demonstrated the benefits of peer-

feedback for the peer-trainee (Falchikov and Goldfinch, 2000). However, the active social engagement in the pre-clinical or clinical tasks also produces cognitive gains for the peer-assessor (Topping, 1996). This occurs as they require an extra effort to ‘understand’ what has happened during the task to be able to feed back to the peer-trainee in simple terms (Ladyshevsky, 2013). In other words, there is a cognitive demand on the peer-assessor to monitor, detect, diagnose and correct her/his peer’s performance that produces the learning benefits (Topping, 2005).

Similarly, the effect on peer-trainees are based on the self-reflection that comes from the non-evaluative dialogic peer-feedback (Ladyshevsky, 2013). The social interaction can therefore transform pre-clinical or clinical experiences into reflective learning which encourages reflective practice (Matheson, 2008). This is an important attribute for dental practitioners who deal with complex professional contexts such as exponential new knowledge and treatment possibilities, more stringent and informed multicultural patients in an increased multidisciplinary focus of practice (Koole *et al.*, 2013). Reflection requires dedicated time (Jonas-Dwyer *et al.*, 2013) which may in part explain its underutilisation as a teaching strategy (Boyd, 2002).

Conclusion

The development of a peer-assessment and peer-feedback protocol has demonstrated its short-term potential to boost students' academic learning and achievement outcomes as well as their reflective skills. However, we would need to follow-up the current students to evaluate the benefits to their professional careers in the medium and long-term.

Future Studies

It is suggested that future studies might involve:

- Evaluation of the effects of the current peer-assessment protocol in other dental student levels such as BDS 3 and BDS 4.
- Assessment of the outcomes of the current protocol implemented in “group practice teams” where students of different BDS levels work together in the treatment of patients. Thus, near peers would assess each other according to their levels of experience.
- Evaluation of the effect of organising “group practice teams” or pairs of students with different levels of reflection skills, thinking that low reflective ones might benefit from high reflective peers.
- Appraisal of the results of implementing the peer-assessment protocol in a multidisciplinary environment where students of different professions work collaboratively together in order to provide a better service to patients. An ideal setting for this might be the outreach “Integrated Team Care Education” programme KCLDI students complete at the University of Portsmouth Dental Academy.
- Replication of the studies in other national or international health-related teaching institutions.
- Repeating the studies in the postgraduate arena. Both within dentistry and other health-related professions.

References

- Adcroft, A. 2011. The Mythology of Feedback. *Higher Education Research & Development*, 30, 405-419.
- Ahmad, N., Naimie, Z., Lui, J., Aziz, A., Abdullah, M., Abu Kasim, N., Abu Kassim, N., Toh, C., Thong, Y. L., Abdul Razak, A., Abdullah, H., Che' Ab. Aziz, Z., Sulaiman, E., Gonzalez, M. A. & Bindal, P. 2012. Clinical Pairing Revisited: A Study at the University of Malaya, Malaysia. *Journal of Dental Education*, 76, 1377-1383.
- Ahmed, K., Miskovic, D., Darzi, A., Athanasiou, T. & Hanna, G. 2011. Observational Tools for Assessment of Procedural Skills: A Systematic Review. *American Journal of Surgery*, 202, 469-480.
- Al-Kadri, H., Al-Kadi, M. & van der Vleuten, C. 2013. Workplace-Based Assessment and Students' Approaches to Learning: A Qualitative Inquiry. *Medical Teacher*, 35, S31-S38.
- Al Ansari, A., Ali, S. & Donnon, T. 2013. The Construct and Criterion Validity of the Mini-Cex: A Meta-Analysis of the Published Research. *Academic Medicine*, 88, 468-474.
- Al Khalifa, K., Al Ansari, A., Violato, C. & Donnon, T. 2013. Multisource Feedback to Assess Surgical Practice: A Systematic Review. *Journal of Surgical Education*, 70, 475-486.
- Albino, J., Inglehart, M. & Tedesco, L. 2012. Dental Education and Changing Oral Health Care Needs: Disparities and Demands. *Journal of Dental Education*, 76, 75-88.
- Albino, J., Young, S., Neumann, L., Kramer, G., Andrieu, S., Henson, L., Horn, B. & Hendricson, W. 2008. Assessing Dental Students' Competence: Best Practice Recommendations in the Performance Assessment Literature and Investigation of Current Practices in Predoctoral Dental Education. *Journal of Dental Education*, 72, 1405-1435.
- Ali, K., Heffernan, E., Lambe, P. & Coombes, L. 2014. Use of Peer Assessment in Tooth Extraction Competency. *European Journal of Dental Education*, 18, 44-50.
- Alves de Lima, A., Conde, D., Costabel, J., Corso, J. & van der Vleuten, C. 2013. A Laboratory Study on the Reliability Estimations of the Mini-Cex. *Advances in Health Sciences Education*, 18, 5-13.
- Ambrose, L. & Ker, J. 2014. Levels of Reflective Thinking and Patient Safety: An

- Investigation of the Mechanisms That Impact on Student Learning in a Single Cohort over a 5 year Curriculum. *Advances in Health Sciences Education*, 19, 297-310.
- American Commission on Dental Accreditation CODA. 2010. *Accreditation Standards for Dental Education Programs* [Online]. Available: <http://www.ada.org/316.aspx>.
- Amulya, J. 2004. What Is Reflective Practice? *The Center for Reflective Community Practice at MIT*, 3, 1-4.
- Archer, J. 2010. State of the Science in Health Professional Education: Effective Feedback. *Medical Education*, 44, 101-108.
- Archer, J. & McAvoy, P. 2011. Factors That Might Undermine the Validity of Patient and Multi-Source Feedback. *Medical Education*, 45, 886-893.
- Archer, J., McGraw, M. & Davies, H. 2010. Assuring Validity of Multisource Feedback in a National Programme. *Archives of Disease in Childhood*, 95, 330-335.
- Archer, J., Norcini, J. & Davies, H. 2005. Use of Sprat for Peer Review of Paediatricians in Training. *British Medical Journal*, 330, 1251-1253.
- Archer, J., Norcini, J., Southgate, L., Heard, S. & Davies, H. 2008. Mini-Pat (Peer Assessment Tool): A Valid Component of a National Assessment Programme in the Uk? *Advances in Health Sciences Education*, 13, 181-192.
- Arnold, L. 2002. Assessing Professional Behavior: Yesterday, Today, and Tomorrow. *Academic Medicine*, 77, 502-515.
- Aronson, L. 2011. Twelve Tips for Teaching Reflection at All Levels of Medical Education. *Medical Teacher*, 33, 200-205.
- Aronson, L., Niehaus, B., Hill-Sakurai, L., Lai, C. & O'Sullivan, P. 2012. A Comparison of Two Methods of Teaching Reflective Ability in Year 3 Medical Students. *Medical Education*, 46, 807-814.
- Aronson, L., Niehaus, B., Lindow, J., Robertson, P. & O'sullivan, P. 2011. Development and Pilot Testing of a Reflective Learning Guide for Medical Education. *Medical Teacher*, 33, e515-e521.
- Asghar, A. 2009. Reciprocal Peer Coaching and Its Use as a Formative Assessment Strategy for First-Year Students. *Assessment & Evaluation in Higher Education*, 35, 403-417.
- Ashley, F., Gibson, B., Daly, B., Lygo Baker, S. & Newton, J. 2006. Undergraduate and Postgraduate Dental Students' Reflection on Learning: A Qualitative Study. *European Journal of Dental Education*, 10, 10-19.
- Audétat, M.-C., Laurin, S., Sanche, G., Béïque, C., Fon, N., Blais, J.-G. & Charlin, B. 2013. Clinical Reasoning Difficulties: A Taxonomy for Clinical Teachers. *Medical*

- Teacher*, 35, e984-e989.
- Aukes, L., Geertsma, J., Cohen-Schotanus, J., Zwierstra, R. & Slaets, J. 2007. The Development of a Scale to Measure Personal Reflection in Medical Practice and Education. *Medical Teacher*, 29, 177-182.
- Australian Dental Council, A. 2010. Professional Attributes and Competencies of the Newly Qualified Dentist. Melbourne, Victoria, Australia.
- Baernstein, A. & Fryer-Edwards, K. 2003. Promoting Reflection on Professionalism: A Comparison Trial of Educational Interventions for Medical Students. *Academic Medicine*, 78, 742-747.
- Bailey, R. & Garner, M. 2010. Is the Feedback in Higher Education Assessment Worth the Paper It Is Written On? Teachers' Reflections on Their Practices. *Teaching in Higher Education*, 15, 187-198.
- Bandura, A. 1997. *Self-Efficacy: The Exercise of Control*, New York, NY, Basingstoke: W. H. Freeman
- Barber, M., Donnelly, K. & Rizvi, S. 2013. *An Avalanche Is Coming: Higher Education and the Revolution Ahead*, London, UK, Institute for Public Policy Research, IPPR.
- Barley, M. 2012. Learning from Reflective Practice and Metacognition – an Anaesthetist's Perspective. *Reflective Practice*, 13, 271-280.
- Barr, H., Freeth, D., Hammick, M., Koppel, I. & Reeves, S. 2000. Evaluations of Interprofessional Education: A United Kingdom Review of Health and Social Care. The United Kingdom Centre for the Advancement of Interprofessional Education (CAIPE) with The British Educational Research Association (BERA).
- Barr, R. & Tagg, J. 1995. From Teaching to Learning: A New Paradigm for Undergraduate Education. *Change*, 27, 12-25.
- Basehore, P., Pomerantz, S. & Gentile, M. 2014. Reliability and Benefits of Medical Student Peers in Rating Complex Clinical Skills. *Medical Teacher*, 36, 409-414.
- Basker, R. 2006. General Dental Council: The First 50 Years.
- Beard, J. 2011. Workplace-Based Assessment: The Need for Continued Evaluation and Refinement. *Surgeon*, 9, S12-S13.
- Beard, J., Marriott, J., Purdie, H. & Crossley, J. 2011. Assessing the Surgical Skills of Trainees in the Operating Theatre: A Prospective Observational Study of the Methodology. *Health Technology Assessment*, 15, 1-168.
- Beard, J., O'sullivan, P., Palmer, B., Qiu, M. & Kim, E. 2012. Peer Assisted Learning in Surgical Skills Laboratory Training: A Pilot Study. *Medical Teacher*, 34, 957-959.
- Beaumont, C., O'Doherty, M. & Shannon, L. 2011. Reconceptualising Assessment

- Feedback: A Key to Improving Student Learning? *Studies in Higher Education*, 36, 671-687.
- Bennett, D., Kelly, M. & O'Flynn, S. 2012. Framework for Feedback: The Peer Mini-Clinical Examination as a Formative Assessment Tool. *Medical Education*, 46, 512.
- Biggs, J. & Tang, C. 2011. *Teaching for Quality Learning at University*, Maidenhead, Berkshire, UK, McGraw-Hill Companies, Incorporated.
- Bindal, N., Goodyear, H., Bindal, T. & Wall, D. 2013. Dops Assessment: A Study to Evaluate the Experience and Opinions of Trainees and Assessors†. *Medical Teacher*, 35, e1230-e1234.
- Bindal, T., Wall, D. & Goodyear, H. 2011. Trainee Doctors' Views on Workplace-Based Assessments: Are They Just a Tick Box Exercise? *Medical Teacher*, 33, 919-927.
- Bing-You, R., Paterson, J. & Levine, M. 1997. Feedback Falling on Deaf Ears: Residents' Receptivity to Feedback Tempered by Sender Credibility. *Medical Teacher*, 19, 40-44.
- Bing-You, R. & Trowbridge, R. 2009. Why Medical Educators May Be Failing at Feedback. *Journal of the American Medical Association*, 302, 1330-1331.
- Black, P. & Wiliam, D. 1998. Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5, 7-74.
- Black, P. & Wiliam, D. 2009. Developing the Theory of Formative Assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31.
- Bleakley, A. 2006. Broadening Conceptions of Learning in Medical Education: The Message from Teamworking. *Medical Education*, 40, 150-157.
- Bloxham, S. 2009. Marking and Moderation in the UK: False Assumptions and Wasted Resources. *Assessment & Evaluation in Higher Education*, 34, 209-220.
- Bloxham, S. & Campbell, L. 2010. Generating Dialogue in Assessment Feedback: Exploring the Use of Interactive Cover Sheets. *Assessment & Evaluation in Higher Education*, 35, 291-300.
- Boehler, M., Rogers, D., Schwind, C., Mayforth, R., Quin, J., Williams, R. & Dunnington, G. 2006. An Investigation of Medical Student Reactions to Feedback: A Randomised Controlled Trial. *Medical Education*, 40, 746-749.
- Boenink, A., Oderwald, A., De Jonge, P., Van Tilburg, W. & Smal, J. 2004. Assessing Student Reflection in Medical Practice. The Development of an Observer-Rated Instrument: Reliability, Validity and Initial Experiences. *Medical Education*, 38, 368-377.

- Bok, H., Teunissen, P., Favier, R., Rietbroek, N., Theyse, L., Brommer, H., Haarhuis, J., van Beukelen, P., van der Vleuten, C. & Jaarsma, D. 2013. Programmatic Assessment of Competency-Based Workplace Learning: When Theory Meets Practice. *BMC Medical Education*, 13, 123.
- Boud, D. 1991. *Implementing Student Self Assessment*, Higher Education Research & Development Society of Australasia Incorporated.
- Boud, D. 1999. Avoiding the Traps: Seeking Good Practice in the Use of Self Assessment and Reflection in Professional Courses. *Social Work Education*, 18, 121-132.
- Boud, D., Cohen, R. & Sampson, J. 1999. Peer Learning and Assessment. *Assessment & Evaluation in Higher Education*, 24, 413-426.
- Boud, D. & Falchikov, N. 2007. *Rethinking Assessment in Higher Education: Learning for the Longer Term* Oxon, UK, Taylor and Francis.
- Boud, D., Keogh, R. & Walker, D. 1985. *Reflection: Turning Experience into Learning*, London, Kogan Page.
- Boud, D. & Molloy, E. 2013. *Feedback in Higher and Professional Education: Understanding It and Doing It Well*, Oxon, UK, Routledge.
- Boud, D. & Walker, D. 1998. Promoting Reflection in Professional Courses: The Challenge of Context. *Studies in Higher Education*, 23, 191-206.
- Boursicot, K., Etheridge, L., Setna, Z., Sturrock, A., Ker, J., Smee, S. & Sambandam, E. 2011. Performance in Assessment: Consensus Statement and Recommendations from the Ottawa Conference. *Medical Teacher*, 33, 370-383.
- Bowen, J. 2006. Medical Education: Educational Strategies to Promote Clinical Diagnostic Reasoning. *New England Journal of Medicine*, 355, 2217-2225.
- Boyd, L. 2002. Reflections on Clinical Practice by First-Year Dental Students: A Qualitative Study. *Journal of Dental Education*, 66, 710-720.
- Boyd, L. 2008. Development of Reflective Judgement in the Pre-Doctoral Dental Clinical Curriculum. *European Journal of Dental Education*, 12, 149-158.
- Branch, W., Jr. & Paranjape, A. 2002. Feedback and Reflection: Teaching Methods for Clinical Settings. *Academic Medicine*, 77, 1185-1188.
- Bridges, S., Chang, J. W., Chu, C. H. & Gardner, K. 2014. Blended Learning in Situated Contexts: 3-Year Evaluation of an Online Peer Review Project. *European Journal of Dental Education*, 18, 170-179.
- Brown, G. & Manogue, M. 2001. Amee Medical Education Guide No. 22: Refreshing Lecturing: A Guide for Lecturers. *Medical Teacher*, 23, 231-244.

- Brown, P., Roediger, H. & McDaniel, M. 2014. *Make It Stick: The Science of Successful Learning*, Harvard, USA, Harvard University Press.
- Brown, S. & Glasner, A. 1999. *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*, Society for Research into Higher Education, Ltd., London (England) / Open University Press.
- Brown, S. & Knight, P. 1998. *Assessing Learners in Higher Education*, London, UK, Kogan Page Limited.
- Bruner, J. 1966. *Toward a Theory of Instruction*, Cambridge, MA, USA, Belknap Press of Harvard University.
- Burgess, A., Roberts, C., Black, K. & Mellis, C. 2013. Senior Medical Student Perceived Ability and Experience in Giving Peer Feedback in Formative Long Case Examinations. *BMC Medical Education*, 13, 79-87.
- Burnard, P. 1995. Nurse Educators' Perceptions of Reflection and Reflective Practice: A Report of a Descriptive Study. *Journal of Advanced Nursing*, 21, 1167-1174.
- Bush, H. & Bissell, V. 2008. The Evaluation of an Approach to Reflective Learning in the Undergraduate Dental Curriculum. *European Journal of Dental Education*, 12, 103-110.
- Cano, F. 2005. Epistemological Beliefs and Approaches to Learning: Their Change through Secondary School and Their Influence on Academic Performance. *British Journal of Educational Psychology*, 75, 203-221.
- Cantillon, P. & Wood, D. 2010. Direct Observation Tools for Workplace-Based Assessment. In: CANTILLON, P. & WOOD, D. (eds.) *Abc of Learning and Teaching in Medicine*. 2nd ed. West Sussex, UK: Wiley-Blackwell.
- Carless, D. 2006. Differing Perceptions in the Feedback Process. *Studies in Higher Education*, 31, 219-233.
- Carless, D. 2007. Learning-Oriented Assessment: Conceptual Bases and Practical Implications. *Innovations in Education and Teaching International*, 44, 57-66.
- Carless, D., Salter, D., Yang, M. & Lam, J. 2010. Developing Sustainable Feedback Practices. *Studies in Higher Education*, 36, 395-407.
- Carmines, E. G. & Zeller, R. A. 1979. *Reliability and Validity Assessment*, London, UK., SAGE Publications Ltd.
- Carr, S. 2006. The Foundation Programme Assessment Tools: An Opportunity to Enhance Feedback to Trainees? *Postgraduate Medical Journal*, 82, 576-579.
- Carr, S. & Johnson, P. 2013. Does Self Reflection and Insight Correlate with Academic Performance in Medical Students? *BMC Medical Education*, 13, 113-117.

- Cartney, P. 2010. Exploring the Use of Peer Assessment as a Vehicle for Closing the Gap between Feedback Given and Feedback Used. *Assessment & Evaluation in Higher Education*, 35, 551-564.
- Chadwick, S., Bearn, D., Jack, A. & O'Brien, K. 2002. Orthodontic Undergraduate Education: Developments in a Modern Curriculum. *European Journal of Dental Education*, 6, 57-63.
- Chadwick, S. & Holsgrove, G. 2009. New Developments in Assessment in Orthodontics. *Journal of Orthodontics*, 36, 122-129.
- Chambers, S., Brosnan, C. & Hassell, A. 2011. Introducing Medical Students to Reflective Practice. *Education for Primary Care*, 22, 100-105.
- Chelliah, K. & Arumugam, Z. 2012. Does Reflective Practice Enhance Clinical Competency in Medical Imaging Undergraduates? *Procedia - Social and Behavioral Sciences*, 60, 73-77.
- Cheng, W. & Warren, M. 1997. Having Second Thoughts: Student Perceptions before and after a Peer Assessment Exercise. *Studies in Higher Education*, 22, 233-239.
- Chickering, A. & Gamson, Z. 1999. Development and Adaptations of the Seven Principles for Good Practice in Undergraduate Education. *New Directions for Teaching and Learning*, 1999, 75-81.
- Cho, K. & MacArthur, C. 2010. Student Revision with Peer and Expert Reviewing. *Learning and Instruction*, 20, 328-338.
- Cho, K., Schunn, C. & Wilson, R. 2006. Validity and Reliability of Scaffolded Peer Assessment of Writing from Instructor and Student Perspectives. *Journal of Educational Psychology*, 98, 891-901.
- Chomsky, N. 1975. *Reflections on Language*, New York, NY, Pantheon Books.
- Clarke, F. 2011. Injecting Expertise: Developing an Expertise-Based Pedagogy for Teaching Local Anaesthesia in Dentistry. *Higher Education Research Network Journal*, 2, 29-43.
- Cleary, T. & Sandars, J. 2011. Assessing Self-Regulatory Processes During Clinical Skill Performance: A Pilot Study. *Medical Teacher*, 33, e368-e374.
- Cobb, K., Brown, G., Jaarsma, D. & Hammond, R. 2013. The Educational Impact of Assessment: A Comparison of Dops and Mcqs. *Medical Teacher*, 35, e1598-e1607.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale, New Jersey, USA, Lawrence Erlbaum Associates.
- Cohen, S., Farrant, P. & Taibjee, S. 2009. Assessing the Assessments: U.K. Dermatology Trainees' Views of the Workplace Assessment Tools. *British Journal of*

- Dermatology*, 161, 34-39.
- Collini, S. 2012. *What Are Universities For?*, London, UK., Penguin Books Ltd.
- Cook, D. & Beckman, T. 2006. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. *The American Journal of Medicine*, 119, 166.e7-166.e16.
- Cook, D., Beckman, T. & Bordage, G. 2007. Quality of Reporting of Experimental Studies in Medical Education: A Systematic Review. *Medical Education*, 41, 737-745.
- Cook, D., Beckman, T., Mandrekar, J. & Pankratz, V. 2010. Internal Structure of Mini-Cex Scores for Internal Medicine Residents: Factor Analysis and Generalizability. *Advances in Health Sciences Education*, 15, 633-645.
- Coomarasamy, A. & Khan, K. 2004. What Is the Evidence That Postgraduate Teaching in Evidence Based Medicine Changes Anything? A Systematic Review. *British Medical Journal*, 329, 1017.
- COPDEND 2006. *A Curriculum for UK Dental Foundation Programme Training*, Department of Health, London.
- COPDEND 2009. *Dental Foundation Training Portfolio & Assessment: User Guide*.
- Corbin, J. & Strauss, A. 1990. Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. *Qualitative Sociology*, 13, 3-21.
- Crisp, B. 2007. Is It Worth the Effort? How Feedback Influences Students' Subsequent Submission of Assessable Work. *Assessment & Evaluation in Higher Education*, 32, 571-581.
- Cronbach, L. 1951. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16, 297-334.
- Cronbach, L. & Meehl, P. 1955. Construct Validity in Psychological Tests. *Psychological Bulletin*, 52, 281-302.
- Cronbach, L. & Shavelson, R. 2004. My Current Thoughts on Coefficient Alpha and Successor Procedures. *Educational and Psychological Measurement*, 64, 391-418.
- Cross, K. 1986. *Adults as Learners: Increasing Participation and Facilitating Learning*, London, UK, Jossey-Bass Limited.
- Crossley, J., Davies, H., Humphris, G. & Jolly, B. 2002a. Generalisability: A Key to Unlock Professional Assessment. *Medical Education*, 36, 972-978.
- Crossley, J., Humphris, G. & Jolly, B. 2002b. Assessing Health Professionals. *Medical Education*, 36, 800-804.
- Crossley, J., Johnson, G., Booth, J. & Wade, W. 2011. Good Questions, Good Answers:

- Construct Alignment Improves the Performance of Workplace-Based Assessment Scales. *Medical Education*, 45, 560-569.
- Crossley, J. & Jolly, B. 2012. Making Sense of Work-Based Assessment: Ask the Right Questions, in the Right Way, About the Right Things, of the Right People. *Medical Education*, 46, 28-37.
- Crossley, J., Russell, J., Jolly, B., Ricketts, C., Roberts, C., Schuwirth, L. & Norcini, J. 2007. 'I'm Pickin' up Good Regressions': The Governance of Generalisability Analyses. *Medical Education*, 41, 926-934.
- Cushing, A., Abbott, S., Lothian, D., Hall, A. & Westwood, O. 2011. Peer Feedback as an Aid to Learning – What Do We Want? Feedback. When Do We Want It? Now! *Medical Teacher*, 33, e105-e112.
- Dale, E. 1946. *Audio- Visual Methods in Teaching*, New York, USA, The Dryden Press.
- Dannefer, E., Henson, L., Bierer, S., Grady-Weliky, T., Meldrum, S., Nofziger, A., Barclay, C. & Epstein, R. 2005. Peer Assessment of Professional Competence. *Medical Education*, 39, 713-722.
- Davies, D. & Jacobs, A. 1985. "Sandwiching" Complex Interpersonal Feedback. *Small Group Behavior*, 16, 387-396.
- Davies, H., Archer, J., Southgate, L. & Norcini, J. 2009. Initial Evaluation of the First Year of the Foundation Assessment Programme. *Medical Education*, 43, 74-81.
- Davies, J., Ciantar, J., Jubraj, B. & Bates, I. 2013. Use of a Multisource Feedback Tool to Develop Pharmacists in a Postgraduate Training Program. *American Journal of Pharmaceutical Education*, 77, 1-7.
- Davis, J. D. 2002. Comparison of Faculty, Peer, Self, and Nurse Assessment of Obstetrics and Gynecology Residents. *Obstetrics and Gynecology*, 99, 647-651.
- Dekker, H., Schonrock-Adema, J., Snoek, J., van der Molen, T. & Cohen-Schotanus, J. 2013. Which Characteristics of Written Feedback Are Perceived as Stimulating Students' Reflective Competence: An Exploratory Study. *BMC Medical Education*, 13, 1-14.
- Delfino, A., Chandratilake, M., Altermatt, F. & Echevarria, G. 2013. Validation and Piloting of Direct Observation of Practical Skills Tool to Assess Intubation in the Chilean Context. *Medical Teacher*, 35, 231-236.
- Delva, D., Sargeant, J., Miller, S., Holland, J., Alexiadis Brown, P., Leblanc, C., Lightfoot, K. & Mann, K. 2013. Encouraging Residents to Seek Feedback. *Medical Teacher*, 35, e1625-e1631.
- Denehy, G. & Fuller, J. 1974. Student Peer Evaluation: An Adjunct to Preclinical

- Laboratory Evaluation. *Journal of Dental Education*, 38, 200-203.
- Dennick, R. 2012. Twelve Tips for Incorporating Educational Theory into Teaching Practices. *Medical Teacher*, 34, 618-624.
- DeVon, H., Block, M., Moyle-Wright, P., Ernst, D. M., Hayden, S., Lazzara, D., Savoy, S. & Kostas-Polston, E. 2007. A Psychometric Toolbox for Testing Validity and Reliability. *Journal of Nursing Scholarship*, 39, 155-164.
- Dewey, J. 1909. *How We Think*, Lexington, MA, USA, D.C. Heath & Company.
- Dewey, J. 1938. *Experience and Education*, New York, NY, USA, Touchstone.
- Dewi, S. & Achmad, T. 2010. Optimising Feedback Using the Mini-Cex During the Final Semester Programme. *Medical Education*, 44, 509.
- Divaris, K., Barlow, P., Chendea, S., Cheong, W., Dounis, A., Dragan, I., Hamlin, J., Hosseinzadeh, L., Kuin, D., Mitirattanakul, S., Mo'nes, M., Molnar, N., Perryer, G., Pickup, J., Raval, N., Shanahan, D., Songpaisan, Y., Taneva, E., Yaghoub-Zadeh, S., West, K. & Vrazic, D. 2008. The Academic Environment: The Students' Perspective. *European Journal of Dental Education*, 12, 120-130.
- Dochy, F., Segers, M. & Sluijsmans, D. 1999. The Use of Self-, Peer and Co-Assessment in Higher Education: A Review. *Studies in Higher Education*, 24, 331-350.
- Dolmans, D. 2013. Self-Assessment and Dialogue: Can It Improve Learning? *Advances in Health Sciences Education*, 18, 193-195.
- Dolmans, D. & Tigelaar, D. 2012. Building Bridges between Theory and Practice in Medical Education Using a Design-Based Research Approach: Amee Guide No. 60. *Medical Teacher*, 34, 1-10.
- Donnon, T., Al Ansari, A., Al Alawi, S. & Violato, C. 2014. The Reliability, Validity, and Feasibility of Multisource Feedback Physician Assessment: A Systematic Review. *Academic Medicine*, 89, 511-516.
- Dornan, T., Carroll, C. & Parboosingh, J. 2002. An Electronic Learning Portfolio for Reflective Continuing Professional Development. *Medical Education*, 36, 767-769.
- Dowden, T., Pittaway, S., Yost, H. & McCarthy, R. 2011. Students' Perceptions of Written Feedback in Teacher Education: Ideally Feedback Is a Continuing Two-Way Communication That Encourages Progress. *Assessment & Evaluation in Higher Education*, 38, 349-362.
- Downing, S. 2003. Validity: On the Meaningful Interpretation of Assessment Data. *Medical Education*, 37, 830-837.
- Dreyfus, H., Dreyfus, S. & Athanasiou, T. 1988. *Mind over Machine: The Power of*

- Human Intuition and Expertise in the Era of the Computer*, New York, USA, Free Press.
- Dreyfus, S. 2004. The Five-Stage Model of Adult Skill Acquisition. *Bulletin of Science, Technology & Society*, 24, 177-181.
- Driessen, E. & Scheele, F. 2013. What Is Wrong with Assessment in Postgraduate Training? Lessons from Clinical Practice and Educational Research. *Medical Teacher*, 35, 569-574.
- Driessen, E., van Tartwijk, J., Govaerts, M., Teunissen, P. & van der Vleuten, C. 2012. The Use of Programmatic Assessment in the Clinical Workplace: A Maastricht Case Report. *Medical Teacher*, 34, 226-231.
- Driessen, E., van Tartwijk, J., Overeem, K., Vermunt, J. & van der Vleuten, C. 2005. Conditions for Successful Reflective Use of Portfolios in Undergraduate Medical Education. *Medical Education*, 39, 1230-1235.
- Durning, S., Artino, A., Boulet, J., La Rochelle, J., van der Vleuten, C., Arze, B. & Schuwirth, L. 2012. The Feasibility, Reliability, and Validity of a Post-Encounter Form for Evaluating Clinical Reasoning. *Medical Teacher*, 34, 30-37.
- Eardley, I., Bussey, M., Woodthorpe, A., Munsch, C. & Beard, J. 2013. Workplace-Based Assessment in Surgical Training: Experiences from the Intercollegiate Surgical Curriculum Programme. *ANZ Journal of Surgery*, 83, 448-53.
- Earl, L. 2012. *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*, Washington DC. USA, SAGE Publications.
- Entwistle, N. & Hounsell, D. 1977. *How Students Learn*, Lancaster, UK, University of Lancaster.
- Epstein, R. 1999. Mindful Practice. *Journal of the American Medical Association*, 282, 833-839.
- Epstein, R. 2007. Assessment in Medical Education. *New England Journal of Medicine*, 356, 387-396.
- Epstein, R. & Hundert, E. 2002. Defining and Assessing Professional Competence. *Journal of the American Medical Association*, 287, 226-235.
- Eraut, M. 1995. Schon Shock: A Case for Refraining Reflection-in-Action? *Teachers and Teaching*, 1, 9-22.
- Ericsson, K. 2008. Deliberate Practice and Acquisition of Expert Performance: A General Overview. *Academic Emergency Medicine*, 15, 988-994.
- Ertmer, P. & Newby, T. 1993. Behaviorism, Cognitivism, Constructivism: Comparing Critical Features from an Instructional Design Perspective. *Performance*

- Improvement Quarterly*, 6, 50-72.
- Escudier, M., Tricio, J. & Odell, E. 2014. Student Acceptability of High-Stakes E-Assessment in Dental Education: Using Privacy Screen Filters to Control Cheating. *Journal of Dental Education*, 78, 558-566.
- Eva, K., Armson, H., Holmboe, E., Lockyer, J., Loney, E., Mann, K. & Sargeant, J. 2012. Factors Influencing Responsiveness to Feedback: On the Interplay between Fear, Confidence, and Reasoning Processes. *Advances in Health Sciences Education*, 17, 15-26.
- Eva, K., Neville, A. J. & Norman, G. R. 1998. Exploring the Etiology of Content Specificity: Factors Influencing Analogic Transfer and Problem Solving. *Academic Medicine*, 73, S1-5.
- Eva, K. & Regehr, G. 2011. Exploring the Divergence between Self-Assessment and Self-Monitoring. *Advances in Health Sciences Education*, 16, 311-329.
- Evans, A., Leeson, R. & Petrie, A. 2007. Reliability of Peer and Self-Assessment Scores Compared with Trainers' Scores Following Third Molar Surgery. *Medical Education*, 41, 866-872.
- Evans, C. 2013. Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research*, 83, 70-120.
- Falchikov, N. 1995. Peer Feedback Marking: Developing Peer Assessment. *Innovations in Education & Training International*, 32, 175-187.
- Falchikov, N. 2007. The Place of Peers in Learning and Assessment. In: BOUD, D. & FALCHIKOV, N. (eds.) *Rethinking Assessment in Higher Education: Learning for the Longer Term*. Oxon, UK: Taylor & Francis.
- Falchikov, N. & Goldfinch, J. 2000. Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*, 70, 287-322.
- Fan, X. 2001. Statistical Significance and Effect Size in Education Research: Two Sides of a Coin. *The Journal of Educational Research*, 94, 275-282.
- Ferguson, J., Wakeling, J. & Bowie, P. 2014. Factors Influencing the Effectiveness of Multisource Feedback in Improving the Professional Practice of Medical Doctors: A Systematic Review. *BMC Medical Education*, 14, 76.
- Finger, M. & Asún, J. 2001. *Adult Education at the Crossroads: Learning Our Way Out*, London, UK, Zed Books.
- Finlay, L. 2008. Reflecting on 'Reflective Practice'. The Open University, UK: Practice-based Professional Learning Centre.

- Finn, G. & Garner, J. 2011. Twelve Tips for Implementing a Successful Peer Assessment. *Medical Teacher*, 33, 443-446.
- Fish, D. & Coles, C. 2005. *Medical Education: Developing a Curriculum for Practice*, Maidenhead, Berkshire, UK, Open University Press.
- Forsyth, D. 2003. *The Professor's Guide to Teaching: Psychological Principles and Practices*, American Psychological Association.
- Foster, K. & Laurent, R. 2013. How We Make Good Doctors into Good Teachers: A Short Course to Support Busy Clinicians to Improve Their Teaching Skills. *Medical Teacher*, 35, 4-7.
- Frankford, D., Patterson, M. & Konrad, T. 2000. Transforming Practice Organizations to Foster Lifelong Learning and Commitment to Medical Professionalism. *Academic Medicine*, 75, 708-717.
- Friedlander, L. & Anderson, V. 2011. A New Predoctoral Endodontic Module: Evaluating Learning and Effectiveness. *Journal of Dental Education*, 75, 351-359.
- Fry, H., Ketteridge, S. & Marshall, S. 2009. *A Handbook for Teaching and Learning in Higher Education: Enhancing Academic Practice*, Oxon, UK, Taylor & Francis.
- Fugill, M. 2013. Defining the Purpose of Phantom Head. *European Journal of Dental Education*, 17, e1-e4.
- General Dental Council 1990. *Recommendations Concerning the Dental Curriculum*, GDC, UK.
- General Dental Council 1997. *The First Five Years: The Undergraduate Dental Curriculum*. London, UK: GDC.
- General Dental Council 2002. *The First Five Years: A Framework for Undergraduate Dental Education*. London, UK: GDC.
- General Dental Council 2008. *The First Five Years*. London, UK: GDC.
- General Dental Council 2012a. *Preparing for Practice: Dental Team Learning Outcomes for Registration*. London, UK: GDC.
- General Dental Council 2012b. *Quality Assurance Process*. London, UK: GDC.
- General Dental Council 2012c. *Standards for Education: Standards and Requirements for Providers of Education and Training Programmes*. London, UK: GDC.
- Gibbs, G. & Simpson, C. 2004. Conditions under Which Assessment Supports Students' Learning. *Learning and Teaching in Higher Education*, 1, 3-31.
- Gillan, C., Lovrics, E., Halpern, E., Wiljer, D. & Harnett, N. 2011. The Evaluation of Learner Outcomes in Interprofessional Continuing Education: A Literature Review and an Analysis of Survey Instruments. *Medical Teacher*, 33, e461-e470.

- Goldie, J. 2013. Assessment of Professionalism: A Consolidation of Current Thinking. *Medical Teacher*, 35, e952-e956.
- Gonzalez, M., Abu Kasim, N. & Naimie, Z. 2013. Soft Skills and Dental Education. *European Journal of Dental Education*, 17, 73-82.
- Gordon, M. & Findley, R. 2011. Educational Interventions to Improve Handover in Health Care: A Systematic Review. *Medical Education*, 45, 1081-1089.
- Gracey, C., Haidet, P., Branch, W., Weissmann, P., Kern, D., Mitchell, G., Frankel, R. & Inui, T. 2005. Precepting Humanism: Strategies for Fostering the Human Dimensions of Care in Ambulatory Settings. *Academic Medicine*, 80, 21-28.
- Grant, A., Kinnersley, P., Metcalf, E., Pill, R. & Houston, H. 2006. Students' Views of Reflective Learning Techniques: An Efficacy Study at a UK Medical School. *Medical Education*, 40, 379-388.
- Grant, J. 2010. Principles of Curriculum Design. In: SWANWICK, T. (ed.) *Workplace Assessment In: Understanding Medical Education: Evidence, Theory and Practice*. 1st ed. West Sussex, UK: Wiley-Blackwell.
- Gregory, R. 2004. *Psychological Testing: History, Principles, and Applications*, Boston, MA, USA, Pearson.
- Grievesson, B., Kirton, J., Palmer, N. & Balmer, M. 2011a. Evaluation of Workplace Based Assessment Tools in Dental Foundation Training. *British Dental Journal*, 211, E8-E8.
- Grievesson, B., Kirton, J. A., Palmer, N. O. A. & Balmer, M. C. 2011b. Evaluation of Workplace Based Assessment Tools in Dental Foundation Training. *British Dental Journal*, 211, 1-5.
- Gukas, I., Miles, S., Heylings, D. & Leinster, S. 2008. Medical Students' Perceptions of Peer Feedback on an Anatomy Student-Selected Study Module. *Medical Teacher*, 30, 812-814.
- Haden, N., Hendricson, W., Kassebaum, D., Ranney, R., Weinstein, G., Anderson, E. & Valachovic, R. 2010. Curriculum Change in Dental Education, 2003–09. *Journal of Dental Education*, 74, 539-557.
- Hafferty, F. & Franks, R. 1994. The Hidden Curriculum, Ethics Teaching, and the Structure of Medical Education. *Academic Medicine*, 69, 861-871.
- Hallett, C. 1997. Learning through Reflection in the Community: The Relevance of Schon's Theories of Coaching to Nursing Education. *International Journal of Nursing Studies*, 34, 103-110.
- Hamer, J., Purchase, H., Luxton-Reilly, A. & Denny, P. 2014. A Comparison of Peer

- and Tutor Feedback. *Assessment & Evaluation in Higher Education*, Early Online.
- Harden, R. & Crosby, J. 2000. Amee Guide No 20: The Good Teacher Is More Than a Lecturer - the Twelve Roles of the Teacher. *Medical Teacher*, 22, 334-347.
- Harden, R. & Gleeson, F. 1979. Assessment of Clinical Competence Using an Objective Structured Clinical Examination (OSCE). *Medical Education*, 13, 39-54.
- Harden, R., Grant, J., Buckley, G. & Hart, I. 1999. Beme Guide No. 1: Best Evidence Medical Education. *Medical Teacher*, 21, 553-562.
- Harden, R. & Laidlaw, J. 2012. *Essential Skills for a Medical Teacher: An Introduction to Teaching and Learning in Medicine*, Elsevier Health Sciences UK.
- Harden, R. & Laidlaw, J. 2013. Be Fair to Students: Four Principles That Lead to More Effective Learning. *Medical Teacher*, 35, 27-31.
- Harden, R., Sowden, S. & Dunn, W. 1984. Educational Strategies in Curriculum Development: The Spices Model. *Medical Education*, 18, 284-297.
- Harden, R., Stevenson, M., Downie, W. & Wilson, G. 1975. Assessment of Clinical Competence Using Objective Structured Examination. *British Medical Journal*, 1, 447-451.
- Hargreaves, J. 2004. So How Do You Feel About That? Assessing Reflective Practice. *Nurse Education Today*, 24, 196-201.
- Harks, B., Rakoczy, K., Hattie, J., Besser, M. & Klieme, E. 2013. The Effects of Feedback on Achievement, Interest and Self-Evaluation: The Role of Feedback's Perceived Usefulness. *Educational Psychology*, 34, 269-290.
- Harlen, W. 2007. *Assessment of Learning*, London, UK, SAGE Publications.
- Hartree, A. 1984. Malcolm Knowles' Theory of Andragogy: A Critique. *International Journal of Lifelong Education*, 3, 203-210.
- Hassell, A., Bullock, A., Whitehouse, A., Wood, L., Jones, P. & Wall, D. 2012. Effect of Rating Scales on Scores Given to Junior Doctors in Multi-Source Feedback. *Postgraduate Medical Journal*, 88, 10-14.
- Hatala, R., Cook, D., Zendejas, B., Hamstra, S. & Brydges, R. 2014. Feedback for Simulation-Based Procedural Skills Training: A Meta-Analysis and Critical Narrative Synthesis. *Advances in Health Sciences Education*, 19, 251-272.
- Hattie, J. 2012. *Visible Learning for Teachers: Maximizing Impact on Learning*, Oxon, UK, Routledge.
- Hattie, J. & Timperley, H. 2007. The Power of Feedback. *Review of Educational Research*, 77, 81-112.
- Hatton, N. & Smith, D. 1995. Reflection in Teacher Education: Towards Definition and

- Implementation. *Teaching and Teacher Education*, 11, 33-49.
- Hauser, A. & Bowen, D. 2009. Primer on Preclinical Instruction and Evaluation. *Journal of Dental Education*, 73, 390-398.
- Hays, R., Davies, H., Beard, J. D., Caldon, L., Farmer, E., Finucane, P., McCrorie, P., Newble, D., Schuwirth, L. & Sibbald, G. 2002. Selecting Performance Assessment Methods for Experienced Physicians. *Medical Education*, 36, 910-917.
- Hendricson, W. 2012. Changes in Educational Methodologies in Predoctoral Dental Education: Finding the Perfect Intersection. *Journal of Dental Education*, 76, 118-141.
- Hendricson, W., Andrieu, S., Chadwick, D., Chmar, J., Cole, J., George, M., Glickman, G., Glover, J., Goldberg, J., Haden, N., Meyerowitz, C., Neumann, L., Pyle, M., Tedesco, L., Valachovic, R., Weaver, R., Winder, R., Young, S. & Kalkwarf, K. 2006. Educational Strategies Associated with Development of Problem-Solving, Critical Thinking, and Self-Directed Learning. *Journal of Dental Education*, 70, 925-936.
- Hendricson, W. & Cohen, P. 1999. Future Directions in Dental School Curriculum, Teaching, and Learning. In: AMERICAN ASSOCIATION OF DENTAL SCHOOLS (ed.) *Leadership for the Future: The Dental School in the University. Center for Educational Policy and Research, American Association of Dental Schools*. Washington, DC, USA.
- Hendricson, W. & Kleffner, J. 1998. Curricular and Instructional Implications of Competency-Based Dental Education. *Journal of Dental Education*, 62, 183-196.
- Higgins, E. 1997. Beyond Pleasure and Pain. *American Psychologist*, 52, 1280-1300.
- Higgins, R., Hartley, P. & Skelton, A. 2002. The Conscientious Consumer: Reconsidering the Role of Assessment Feedback in Student Learning. *Studies in Higher Education*, 27, 53-64.
- Higher Education Funding Council for England 2011. The National Student Survey: Findings and Trends 2006–2010. Bristol: Higher Education Funding Council for England.
- Hill, F., Kendall, K., Galbraith, K. & Crossley, J. 2009. Implementing the Undergraduate Mini-Cex: A Tailored Approach at Southampton University. *Medical Education*, 43, 326-334.
- Hobbs, V. 2007. Faking It or Hating It: Can Reflective Practice Be Forced? *Reflective Practice*, 8, 405-417.
- Hoffman, K. & Donaldson, J. 2004. Contextual Tensions of the Clinical Environment

- and Their Influence on Teaching and Learning. *Medical Education*, 38, 448-454.
- Hojat, M. & Xu, G. 2004. A Visitor's Guide to Effect Sizes – Statistical Significance Versus Practical (Clinical) Importance of Research Findings. *Advances in Health Sciences Education*, 9, 241-249.
- Hounsell, D., Blair, S., Falchikov, N., Hounsell, J., Huxham, M., Klampfleitner, M. & Thompson, K. 2007. Innovative Assessment across the Disciplines: An Analytical Review of the Literature. Heslington, United Kingdom: Higher Education Academy.
- Hurst, Y., Prescott-Clements, L. & Rennie, J. 2004. The Patient Assessment Questionnaire: A New Instrument for Evaluating the Interpersonal Skills of Vocational Dental Practitioners. *British Dental Journal*, 197, 497-500.
- Ivers, N., Jamtvedt, G., Flottorp, S., Young Jane, M., Odgaard-Jensen, J., Simon, F., O'Brien Mary, A., Johansen, M., Grimshaw, J. & Oxman, A. 2012. Audit and Feedback: Effects on Professional Practice and Healthcare Outcomes. *Cochrane Database of Systematic Reviews* [Online]. Available: <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD000259.pub3/abstract>.
- Jackson, L. 2009. Revisiting Adult Learning Theory through the Lens of an Adult Learner. *Adult Learning*, 20, 20-22.
- Jacobs, R., Briggs, D. & Whitney, D. 1975. Continuous-Progress Education: Iii. Student Self-Evaluation and Peer Evaluation. *Journal of Dental Education*, 39, 535-541.
- Jahangiri, L. & Mucciolo, T. 2011. Toward a Model of Institutional Scholarship in Health Professions Education. *Journal of Dental Education*, 75, 1569-1576.
- Jarvis, P. 1987. *Adult Learning in the Social Context*, Kent, UK, Croom Helm Ltd.
- Jonas-Dwyer, D., Abbott, P. & Boyd, N. 2013. First Reflections: Third-Year Dentistry Students' Introduction to Reflective Practice. *European Journal of Dental Education*, 17, e64-e69.
- Kalsi, H., Kalsi, J. & Fisher, N. 2013. An Explanation of Workplace-Based Assessments in Postgraduate Dental Training and a Review of the Current Literature. *British Dental Journal*, 215, 519-524.
- Kamp, R., Dolmans, D., Berkel, H. & Schmidt, H. 2013. The Effect of Midterm Peer Feedback on Student Functioning in Problem-Based Tutorials. *Advances in Health Sciences Education*, 18, 199-213.
- Kardos, R., Cook, J., Butson, R. & Kardos, T. 2009. The Development of an Eportfolio for Life-Long Reflective Learning and Auditable Professional Certification. *European Journal of Dental Education*, 13, 135-141.
- Karl, M., Graef, F., Wichmann, M. & Beck, N. 2011. Evaluation of Tooth Preparations

- a Comparative Study between Faculty Members and Pre-Clinical Students. *European Journal of Dental Education*, 15, 250-254.
- Kaufman, D. 2003. Abc of Learning and Teaching in Medicine: Applying Educational Theory in Practice. *British Medical Journal*, 326, 213-216.
- Kaufman, D. 2010. Applying Educational Theory to Practice. In: CANTILLON, P. & WOOD, D. (eds.) *Abc of Learning and Teaching in Medicine*. 2nd. ed. West Sussex, UK: Wiley-Blackwell.
- Kaufman, D. & Mann, K. 2010. Teaching and Learning in Medical Education: How Theory Can Inform Practice. In: SWANWICK, T. (ed.) *Workplace Assessment In: Understanding Medical Education: Evidence, Theory and Practice*. 1st ed. West Sussex, UK: Wiley-Blackwell.
- Kember, D. 1999. Determining the Level of Reflective Thinking from Students' Written Journals Using a Coding Scheme Based on the Work of Mezirow. *International Journal of Lifelong Education*, 18, 18-30.
- Kember, D., Leung, D., Jones, A., Loke, A., McKay, J., Sinclair, K., Tse, H., Webb, C., Yuet Wong, F., Wong, M. & Yeung, E. 2000. Development of a Questionnaire to Measure the Level of Reflective Thinking. *Assessment & Evaluation in Higher Education*, 25, 381-395.
- Kember, D., McKay, J., Sinclair, K. & Wong, F. 2008. A Four-Category Scheme for Coding and Assessing the Level of Reflection in Written Work. *Assessment & Evaluation in Higher Education*, 33, 369-379.
- Ker, J. & Bradley, P. 2011. Simulation in Medical Education. In: SWANWICK, T. (ed.) *Understanding Medical Education: Evidence, Theory and Practice*. 1st ed. West Sussex, UK: Wiley-Blackwell.
- Kern, D., Thomas, P. & Hughes, M. 2009. *Curriculum Development for Medical Education: A Six-Step Approach*, Johns Hopkins University Press.
- Kersten, H. 1997. The Enhancement of Learning by Teaching. *European Journal of Dental Education*, 1, 149-152.
- Kilminster, S., Cottrell, D., Grant, J. & Jolly, B. 2007. Amee Guide No. 27: Effective Educational and Clinical Supervision. *Medical Teacher*, 29, 2-19.
- King's College London Academic Services. 2011. *Kcl Learning Teaching and Assessment Strategy* [Online]. Available: <http://www.kcl.ac.uk/college/policyzone/index.php?id=288>.
- King's College London Dental Institute. 2012. *Bds Proramme Description* [Online]. KCLDI. Available:

- http://www.kcl.ac.uk/prospectus/undergraduate/index/name/dentistry/alpha/D/header_search/ Last accessed March 2014].
- Kings College London Dental Institute, K. 2013. *Dentistry (Bds) Programme Description* [Online]. Available: http://www.kcl.ac.uk/prospectus/undergraduate/details/name/dentistry/alpha/d/header_search/ [Accessed 8th of October.
- Kirkpatrick, D. 1967. Evaluation and Training. In: CRAIG, R. & MITTEL, I. (eds.) *Training and Development Handbook*. New York, USA: McGraw-Hill.
- Kirschner, P., Sweller, J. & Clark, R. 2006. Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41, 75-86.
- Kirton, J., Palmer, N., Grieveson, B. & Balmer, M. 2013. A National Evaluation of Workplace-Based Assessment Tools (Wpbas) in Foundation Dental Training: A UK Study. Effective and Useful but Do They Provide an Equitable Training Experience? *British Dental Journal*, 214, 305-309.
- Kluger, A. & DeNisi, A. 1996. The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, 119, 254-284.
- Kluger, A. & Van Dijk, D. 2010. Feedback, the Various Tasks of the Doctor, and the Feedforward Alternative. *Medical Education*, 44, 1166-1174.
- Kneebone, R. 2002. Total Internal Reflection: An Essay on Paradigms. *Medical Education*, 36, 514-518.
- Knight, P. 2006. The Local Practices of Assessment. *Assessment & Evaluation in Higher Education*, 31, 435-452.
- Knight, P. 2007. Grading, Classifying and Future Learning. In: BOUD, D. & FALCHIKOV, N. (eds.) *Rethinking Assessment in Higher Education: Learning for the Longer Term* 1ed. Oxon, UK: Taylor and Francis.
- Knight, P. 2012. *Assessment for Learning in Higher Education*, London, UK, Kogan Page Limited.
- Knowles, M. 1980. *The Modern Practice of Adult Education: From Pedagogy to Andragogy*, New York, USA, Cambridge Books, the Adult Education Company.
- Knowles, M. 1984. *Andragogy in Action: Applying Modern Principles of Adult Learning*, San Francisco, CA, USA, Jossey-Bass.
- Knowles, M., Holton, E. & Swanson, R. 2012. *The Adult Learner: The Definitive*

- Classic in Adult Education and Human Resource Development*, Oxford, UK, Taylor & Francis.
- Kogan, J., Bellini, L. & Shea, J. 2003. Feasibility, Reliability, and Validity of the Mini-Clinical Evaluation Exercise (Mcex) in a Medicine Core Clerkship. *Academic Medicine*, 78, S33-S35.
- Kogan, J., Conforti, L., Bernabeo, E., Durning, S., Hauer, K. & Holmboe, E. 2012. Faculty Staff Perceptions of Feedback to Residents after Direct Observation of Clinical Skills. *Medical Education*, 46, 201-215.
- Kogan, J. & Hauer, K. 2006. Brief Report: Use of the Mini-Clinical Evaluation Exercise in Internal Medicine Core Clerkships. *Journal of General Internal Medicine*, 21, 501-502.
- Kogan, J., Holmboe, E. & Hauer, K. 2009. Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees: A Systematic Review. *Journal of the American Medical Association*, 302, 1316-1326.
- Kolb, D. 1976. Management and the Learning Process. *California Management Review*, 18, 652-673.
- Kolb, D. 1984. *Experiential Learning: Experience as the Source of Learning and Development*, Ney Jersey, USA, Prentice-Hall.
- Koole, S., Vanobbergen, J., De Visschere, L., Aper, L., Dornan, T. & Derese, A. 2013. The Influence of Reflection on Portfolio Learning in Undergraduate Dental Education. *European Journal of Dental Education*, 17, e93-e99.
- Kossioni, A., Varela, R., Ekonomu, I., Lyrakos, G. & Dimoliatis, I. 2012. Students' Perceptions of the Educational Environment in a Greek Dental School, as Measured by Dreem. *European Journal of Dental Education*, 16, e73-e78.
- Kramer, G., Albino, J., Andrieu, S., Hendricson, W., Henson, L., Horn, B., Neumann, L. & Young, S. 2009. Dental Student Assessment Toolbox. *Journal of Dental Education*, 73, 12-35.
- Ladyshevsky, R. 2010. Building Competency in the Novice Allied Health Professional through Peer Coaching. *Journal of Allied Health*, 39, 77E-82E.
- Ladyshevsky, R. 2013. The Role of Peers in Feedback Processes. In: BOUD, D. & MOLLOY, E. (eds.) *Feedback in Higher and Professional Education: Understanding It and Doing It Well*. Oxon, UK: Routledge.
- Lanning, S., Brickhouse, T., Gunsolley, J., Ranson, S. & Willett, R. 2011. Communication Skills Instruction: An Analysis of Self, Peer-Group, Student Instructors and Faculty Assessment. *Patient Education and Counseling*, 83, 145-151.

- Larsen, T. & Jeppe-Jensen, D. 2008. The Introduction and Perception of an OSCE with an Element of Self- and Peer-Assessment. *European Journal of Dental Education*, 12, 2-7.
- Lave, J. & Wenger, E. 1991. *Situated Learning: Legitimate Peripheral Participation*, Cambridge, UK, Cambridge University Press.
- Learman, L., Autry, A. & O'Sullivan, P. 2008. Reliability and Validity of Reflection Exercises for Obstetrics and Gynecology Residents. *American Journal of Obstetrics and Gynecology*, 198, 461.e1-461.e10.
- Lee, P. & Caffarella, R. 1994. Methods and Techniques for Engaging Learners in Experiential Learning Activities. *New Directions for Adult and Continuing Education*, 1994, 43-54.
- Lethbridge, K., Andrusyszyn, M.-A., Iwasiw, C., Laschinger, H. & Fernando, R. 2011. Assessing the Psychometric Properties of Kember and Leung's Reflection Questionnaire. *Assessment & Evaluation in Higher Education*, 38, 1-23.
- Leung, D. & Kember, D. 2003. The Relationship between Approaches to Learning and Reflection Upon Practice. *Educational Psychology*, 23, 61-71.
- Lew, M. & Schmidt, H. 2011. Self-Reflection and Academic Performance: Is There a Relationship? *Advances in Health Sciences Education*, 16, 529-545.
- Liao, K.-C., Pu, S.-J., Liu, M.-S., Yang, C.-W. & Kuo, H.-P. 2013. Development and Implementation of a Mini-Clinical Evaluation Exercise (Mini-Cex) Program to Assess the Clinical Competencies of Internal Medicine Residents: From Faculty Development to Curriculum Evaluation. *BMC Medical Education*, 13, 1-7.
- Light, G., Cox, R. & Calkins, S. 2009. *Learning and Teaching in Higher Education: The Reflective Professional*, London, UK, SAGE Publications.
- Lim, L.-A. 2011. A Comparison of Students' Reflective Thinking across Different Years in a Problem-Based Learning Environment. *Instructional Science*, 39, 171-188.
- Lindström, U., Hamberg, K. & Johansson, E. 2011. Medical Students' Experiences of Shame in Professional Enculturation. *Medical Education*, 45, 1016-1024.
- Liu, N.-F. & Carless, D. 2006. Peer Feedback: The Learning Element of Peer Assessment. *Teaching in Higher Education*, 11, 279-290.
- Lonka, K., Slotte, V., Halttunen, M., Kurki, T., Tiitinen, A., Vaara, L. & Paavonen, J. 2001. Portfolios as a Learning Tool in Obstetrics and Gynaecology Undergraduate Training. *Medical Education*, 35, 1125-1130.
- Lurie, S., Nofziger, A., Meldrum, S., Mooney, C. & Epstein, R. 2006. Temporal and

- Group-Related Trends in Peer Assessment Amongst Medical Students. *Medical Education*, 40, 840-847.
- Maatsch, J., Huang, R., Downing, S. & Barker, B. 1983. Predictive Validity of Medical Specialist Examinations. National Center of Health Services Research. Office of Medical Education Research and Development: Michigan State University, East Lansing, MI., USA.
- Mackillop, L., Crossley, J., Vivekananda-Schmidt, P., Wade, W. & Armitage, M. 2011a. A Single Generic Multi-Source Feedback Tool for Revalidation of All UK Career-Grade Doctors: Does One Size Fit All? *Medical Teacher*, 33, e75-e83.
- Mackillop, L., Parker-Swift, J. & Crossley, J. 2011b. Getting the Questions Right: Non-Compound Questions Are More Reliable Than Compound Questions on Matched Multi-Source Feedback Instruments. *Medical Education*, 45, 843-848.
- Mamede, S. & Schmidt, H. 2004. The Structure of Reflective Practice in Medicine. *Medical Education*, 38, 1302-1308.
- Mamede, S., Schmidt, H., Rikers, R., Penaforte, J. & Coelho-Filho, J. 2007. Breaking Down Automaticity: Case Ambiguity and the Shift to Reflective Approaches in Clinical Reasoning. *Medical Education*, 41, 1185-1192.
- Mann, K. 2011. Theoretical Perspectives in Medical Education: Past Experience and Future Possibilities. *Medical Education*, 45, 60-68.
- Mann, K., Gordon, J. & MacLeod, A. 2009. Reflection and Reflective Practice in Health Professions Education: A Systematic Review. *Advances in Health Sciences Education*, 14, 595-621.
- Mann, K., van der Vleuten, C., Eva, K., Armson, H., Chesluk, B., Dornan, T., Holmboe, E., Lockyer, J., Loney, E. & Sargeant, J. 2011. Tensions in Informed Self-Assessment: How the Desire for Feedback and Reticence to Collect and Use It Can Conflict. *Academic Medicine*, 86, 1120-1127.
- Manogue, M. & Brown, G. 2007. Managing the Curriculum – for a Change. *European Journal of Dental Education*, 11, 75-86.
- Manogue, M., Brown, G. & Foster, H. 2001. Clinical Assessment of Dental Students: Values and Practices of Teachers in Restorative Dentistry. *Medical Education*, 35, 364-370.
- Manogue, M., Kelly, M., Bartakova Masaryk, S., Brown, G., Catalanotto, F., Choo-Soo, T., Delap, E., Godoroja, P., Morio, I., Rotgans, J. & Saag, M. 2002. 2.1 Evolving Methods of Assessment. *European Journal of Dental Education*, 6, 53-66.
- Manogue, M., McLoughlin, J., Christersson, C., Delap, E., Lindh, C., Schoonheim-

- Klein, M. & Plasschaert, A. 2011. Curriculum Structure, Content, Learning and Assessment in European Undergraduate Dental Education - Update 2010. *European Journal of Dental Education*, 15, 133-141.
- Martineau, B., Mamede, S., St-Onge, C., Rikers, R. & Schmidt, H. 2013. To Observe or Not to Observe Peers When Learning Physical Examination Skills; That Is the Question. *BMC Medical Education*, 13, 1-11.
- Marton, F. & Booth, S. 1997. *Learning and Awareness*, Mahwah, New Jersey, USA, Lawrence Erlbaum Associates.
- Matheson, D. 2008. *An Introduction to the Study of Education*, Oxon, UK, Routledge.
- Mattheos, N., Stefanovic, N., Apse, P., Attstrom, R., Buchanan, J., Brown, P., Camilleri, A., Care, R., Fabrikant, E., Gundersen, S., Honkala, S., Johnson, L., Jonas, I., Kavadella, A., Moreira, J., Peroz, I., Perryer, D., Seemann, R., Tansy, M., Thomas, H., Tsuruta, J., Uribe, S., Urtane, I., Walsh, T., Zimmerman, J. & Walmsley, A. 2008. Potential of Information Technology in Dental Education. *European Journal of Dental Education*, 12, 85-92.
- Mattheos, N., Ucer, C., Van de Velde, T. & Nattestad, A. 2009. Assessment of Knowledge and Competencies Related to Implant Dentistry in Undergraduate and Postgraduate University Education. *European Journal of Dental Education*, 13, 55-65.
- McCormack, W., Lazarus, C., Stern, D. & Small, P., Jr 2007. Peer Nomination: A Tool for Identifying Medical Student Exemplars in Clinical Competence and Caring, Evaluated at Three Medical Schools. *Academic Medicine*, 82, 1033-1039.
- McCoubrie, P. 2004. Improving the Fairness of Multiple-Choice Questions: A Literature Review. *Medical Teacher*, 26, 709-712.
- McDonald, B. 2010. Improving Learning through Meta Assessment. *Active Learning in Higher Education*, 11, 119-129.
- McDowell, L., Wakelin, D., Montgomery, C. & King, S. 2010. Does Assessment for Learning Make a Difference? The Development of a Questionnaire to Explore the Student Response. *Assessment & Evaluation in Higher Education*, 36, 749-765.
- McKenzie, C. 2013. Dental Student Perceptions of Case-Based Educational Effectiveness. *Journal of Dental Education*, 77, 688-694.
- McLeod, P., Brawer, J., Steinert, Y., Chalk, C. & McLeod, A. 2008. A Pilot Study Designed to Acquaint Medical Educators with Basic Pedagogic Principles. *Medical Teacher*, 30, 92-93.
- McMillan, W. 2011. Making the Most of Teaching at the Chairside. *European Journal*

- of *Dental Education*, 15, 63-68.
- Mehta, F., Brown, J. & Shaw, N. 2013. Do Trainees Value Feedback in Case-Based Discussion Assessments? *Medical Teacher*, 35, e1166-e1172.
- Melnick, D., Asch, D., Blackmore, D., Klass, D. & Norcini, J. 2002. Conceptual Challenges in Tailoring Physician Performance Assessment to Individual Practice. *Medical Education*, 36, 931-935.
- Merriam, S. 2001. Andragogy and Self-Directed Learning: Pillars of Adult Learning Theory. *New Directions for Adult and Continuing Education*, 2001, 3-14.
- Merriam, S., Caffarella, R. & Baumgartner, L. 2007. *Learning in Adulthood: A Comprehensive Guide*, San Francisco, CA, USA, Wiley.
- Messick, S. 1995. Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning. *American Psychologist*, 50, 741-749.
- Mezirow, J. 1991. *Transformative Dimensions of Adult Learning*, San Francisco, CA., Jossey-Bass.
- Miettinen, R. 2000. The Concept of Experiential Learning and John Dewey's Theory of Reflective Thought and Action. *International Journal of Lifelong Education*, 19, 54-72.
- Miller, A. & Archer, J. 2010. Impact of Workplace Based Assessment on Doctors' Education and Performance: A Systematic Review. *British Medical Journal*, 341, c5064-c5064.
- Miller, B., Moore, D., Jr., Stead, W. & Balser, J. 2010. Beyond Flexner: A New Model for Continuous Learning in the Health Professions. *Academic Medicine*, 85, 266-272.
- Miller, G. 1961. *Teaching and Learning in Medical School*, Cambridge, Mas, USA, Harvard University Press.
- Miller, G. 1990. The Assessment of Clinical Skills/Competence/Performance. *Academic Medicine*, 65, 63-67.
- Millett, D. 2011. Wbas in Orthodontic Postgraduate Education: Why and How? *Journal of Orthodontics*, 38, 240-242.
- Mitchell, C., Bhat, S., Herbert, A. & Baker, P. 2011. Workplace-Based Assessments of Junior Doctors: Do Scores Predict Training Difficulties? *Medical Education*, 45, 1190-1198.
- Mitchell, C., Bhat, S., Herbert, A. & Baker, P. 2013. Workplace-Based Assessments in Foundation Programme Training: Do Trainees in Difficulty Use Them Differently? *Medical Education*, 47, 292-300.

- Mitchell, R. 1994. The Development of the Cognitive Behavior Survey to Assess Medical Student Learning. *Teaching and Learning in Medicine*, 6, 161-167.
- Mitchell, R., Regan-Smith, M., Fisher, M., Knox, I. & Lambert, D. 2009. A New Measure of the Cognitive, Metacognitive, and Experiential Aspects of Residents' Learning. *Academic Medicine*, 84, 918-926.
- Mofidi, M., Strauss, R., Pitner, L. & Sandler, E. 2003. Dental Students' Reflections on Their Community-Based Experiences: The Use of Critical Incidents. *Journal of Dental Education*, 67, 515-523.
- Moon, J. 2007. Getting the Measure of Reflection: Considering Matters of Definition and Depth. *Journal of Radiotherapy in Practice*, 6, 191-200.
- Moon, J. 2013. *A Handbook of Reflective and Experiential Learning: Theory and Practice*, Oxon, UK, RoutledgeFalmer.
- Moonen-van Loon, J., Overeem, K., Donkers, H., van der Vleuten, C. & Driessen, E. 2013. Composite Reliability of a Workplace-Based Assessment Toolbox for Postgraduate Medical Education. *Advances in Health Sciences Education*, 18, 1087-1102.
- Moore, T. & Kain, D. 2011. Student Tutors for Problem-Based Learning in Dental Hygiene: A Study of Tutor Actions. *Journal of Dental Education*, 75, 805-816.
- Mori, B., Batty, H. & Brooks, D. 2008. The Feasibility of an Electronic Reflective Practice Exercise among Physiotherapy Students. *Medical Teacher*, 30, e232-e238.
- Morris, C. & Blaney, D. 2011. Work-Base Learning. In: SWANWICK, T. (ed.) *Understanding Medical Education: Evidence, Theory and Practice*. 1st ed. West Sussex, UK: Wiley-Blackwell.
- Murray, J., Benington, I., Cadden, S., Irvine, M. & Toppin, D. 2003. Visitation of the Bds Programme of Guy's, King's and St Thomas' Dental Institute. General Dental Council.
- Mutch, A. 2003. Exploring the Practice of Feedback to Students. *Active Learning in Higher Education*, 4, 24-38.
- Naghdi-pour, B. & Emeagwali, O. 2013. Assessing the Level of Reflective Thinking in Elt Students. *Procedia - Social and Behavioral Sciences*, 83, 266-271.
- National Postsecondary Education Cooperative 2000. The Npec Sourcebook on Assessment. Volume 1: Definitions and Assessment Methods for Critical Thinking, Problem Solving, and Writing. In: DEPARTMENT OF EDUCATION NATIONAL CENTER FOR EDUCATION STATISTICS (ed.). Washington, DC: U.S.
- National Student Survey, N. 2014. *The National Student Survey* [Online]. Available:

- <http://www.thestudentsurvey.com/> [Accessed May 2014].
- Neher, J., Gordon, K., Meyer, B. & Stevens, N. 1992. A Five-Step “Microskills” Model of Clinical Teaching. *The Journal of the American Board of Family Practice*, 5, 419-424.
- Nelson, A., Nelson, S., Linn, A., Raw, L., Kildea, H. & Tonkin, A. 2013. Tomorrow's Educators ... Today? Implementing near-Peer Teaching for Medical Students. *Medical Teacher*, 35, 156-159.
- Nesbitt, A., Baird, F., Canning, B., Griffin, A. & Sturrock, A. 2013. Student Perception of Workplace-Based Assessment. *The Clinical Teacher*, 10, 399-404.
- Newble, D. & Cannon, R. 2001. *A Handbook for Medical Teachers*, London, UK, Springer.
- Nicol, D. 2010. From Monologue to Dialogue: Improving Written Feedback Processes in Mass Higher Education. *Assessment & Evaluation in Higher Education*, 35, 501-517.
- Nicol, D. & Macfarlane-Dick, D. 2006. Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Studies in Higher Education*, 31, 199-218.
- Nicol, D., Thomson, A. & Breslin, C. 2014. Rethinking Feedback Practices in Higher Education: A Peer Review Perspective. *Assessment & Evaluation in Higher Education*, 39, 102-122.
- Niemi, P. 1997. Medical Students' Professional Identity: Self-Reflection During the Preclinical Years. *Medical Education*, 31, 408-415.
- Nilson, L. 2003. Improving Student Peer Feedback. *College Teaching*, 51, 34-38.
- Nofziger, A., Naumburg, E., Davis, B., Mooney, C. & Epstein, R. 2010. Impact of Peer Assessment on the Professional Development of Medical Students: A Qualitative Study. *Academic Medicine*, 85, 140-147.
- Norcini, J. 2003a. Peer Assessment of Competence. *Medical Education*, 37, 539-543.
- Norcini, J. 2005. Current Perspectives in Assessment: The Assessment of Performance at Work. *Medical Education*, 39, 880-889.
- Norcini, J. 2011. Workplace Assessment. In: SWANWICK, T. (ed.) *Understanding Medical Education: Evidence, Theory and Practice*. 1st ed. West Sussex, UK: Wiley-Blackwell.
- Norcini, J., Anderson, B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., Galbraith, R., Hays, R., Kent, A., Perrott, V. & Roberts, T. 2011. Criteria for Good Assessment: Consensus Statement and Recommendations from the Ottawa 2010 Conference.

- Medical Teacher*, 33, 206-214.
- Norcini, J., Blank, L., Arnold, G. & Kimball, H. 1995. The Mini-Cex (Clinical Evaluation Exercise): A Preliminary Investigation. *Annals of Internal Medicine*, 123, 795-799.
- Norcini, J., Blank, L., Arnold, G. & Kimball, H. 1997. Examiner Differences in the Mini-Cex. *Advances in Health Sciences Education*, 2, 27-33.
- Norcini, J., Blank, L., Duffy, F. & Fortna, G. 2003. The Mini-Cex: A Method for Assessing Clinical Skills. *Annals of Internal Medicine*, 138, 476-481.
- Norcini, J. & Burch, V. 2007. Workplace-Based Assessment as an Educational Tool: Amee Guide No. 31. *Medical Teacher*, 29, 855-871.
- Norcini, J. & McKinley, D. 2007. Assessment Methods in Medical Education. *Teaching and Teacher Education*, 23, 239-250.
- Norcini, J. J. 2003b. Peer Assessment of Competence. *Medical Education*, 37, 539-543.
- Norman, G. 1999. The Adult Learner: A Mythical Species. *Academic Medicine*, 74, 886-889.
- Norman, G. & Eva, K. 2010. Quantitative Research Methods in Medical Education. In: SWANWICK, T. (ed.) *Understanding Medical Education: Evidence, Theory and Practice*. 1st ed. West Sussex, UK: Wiley-Blackwell.
- Norman, G. & Schmidt, H. 1992. The Psychological Basis of Problem-Based Learning: A Review of the Evidence. *Academic Medicine*, 67, 557-565.
- Norman, G., van der Vleuten, C. & De Graaff, E. 1991. Pitfalls in the Pursuit of Objectivity: Issues of Validity, Efficiency and Acceptability. *Medical Education*, 25, 119-126.
- Nulty, D. 2010. Peer and Self-Assessment in the First Year of University. *Assessment & Evaluation in Higher Education*, 36, 493-507.
- Nulty, D., Short, L. & Johnson, N. 2010. Improving Assessment in Dental Education through a Paradigm of Comprehensive Care: A Case Report. *Journal of Dental Education*, 74, 1367-1379.
- O'Donnell, J., Oakley, M., Haney, S., O'Neill, P. & Taylor, D. 2011. Rubrics 101: A Primer for Rubric Development in Dental Education. *Journal of Dental Education*, 75, 1163-1175.
- Ogden, G., Green, M. & Ker, J. 2000. The Use of Interprofessional Peer Examiners in an Objective Structured Clinical Examination: Can Dental Students Act as Examiners? *British Dental Journal*, 189, 160-164.
- Oliver, R., Kersten, H., Vinkka-Puhakka, H., Alpasan, G., Bearn, D., Cema, I., Delap,

- E., Dummer, P., Goulet, J., Gugushe, T., Jeniati, E., Jerolimov, V., Kotsanos, N., Krifka, S., Levy, G., Neway, M., Ogawa, T., Saag, M., Sidlauskas, A., Skaleric, U., Vervoorn, M. & White, D. 2008. Curriculum Structure: Principles and Strategy. *European Journal of Dental Education*, 12, 74-84.
- Oosterbaan, A., van der Schaaf, M., Baartman, L. & Stokking, K. 2010. Reflection During Portfolio-Based Conversations. *International Journal of Educational Research*, 49, 151-160.
- Ormrod, J. 2011. *Human Learning*, Essex, UK, Pearson College Division.
- Orsmond, P., Maw, S., Park, J., Gomez, S. & Crook, A. 2013. Moving Feedback Forward: Theory to Practice. *Assessment & Evaluation in Higher Education*, 38, 240-252.
- Orsmond, P. & Merry, S. 2011. Feedback Alignment: Effective and Ineffective Links between Tutors' and Students' Understanding of Coursework Feedback. *Assessment & Evaluation in Higher Education*, 36, 125-136.
- Orsmond, P., Merry, S. & Reiling, K. 1996. The Importance of Marking Criteria in the Use of Peer Assessment. *Assessment & Evaluation in Higher Education*, 21, 239-250.
- Orsmond, P., Merry, S. & Reiling, K. 2005. Biology Students' Utilization of Tutors' Formative Feedback: A Qualitative Interview Study. *Assessment & Evaluation in Higher Education*, 30, 369-386.
- Ostapczuk, M., Hugger, A., de Bruin, J., Ritz-Timme, S. & Rotthoff, T. 2012. Dreem on, Dentists! Students' Perceptions of the Educational Environment in a German Dental School as Measured by the Dundee Ready Education Environment Measure. *European Journal of Dental Education*, 16, 67-77.
- Overeem, K., Lombarts, M., Arah, O., Klazinga, N., Grol, R. & Wollersheim, H. 2010. Three Methods of Multi-Source Feedback Compared: A Plea for Narrative Comments and Coworkers' Perspectives. *Medical Teacher*, 32, 141-147.
- Pangaro, L. & ten Cate, O. 2013. Frameworks for Learner Assessment in Medicine: Ameer Guide No. 78. *Medical Teacher*, 35, e1197-e1210.
- Papinczak, T., Young, L. & Groves, M. 2007. Peer Assessment in Problem-Based Learning: A Qualitative Study. *Advances in Health Sciences Education*, 12, 169-186.
- Parkes, J., Abercrombie, S. & McCarty, T. 2013. Feedback Sandwiches Affect Perceptions but Not Performance. *Advances in Health Sciences Education*, 18, 397-407.
- Pavlov, I. 1927. *Conditioned Reflexes*, Oxford, UK, Oxford University Press.

- Pearson, D. & Heywood, P. 2004. Portfolio Use in General Practice Vocational Training: A Survey of Gp Registrars. *Medical Education*, 38, 87-95.
- Pee, B., Woodman, T., Fry, H. & Davenport, E. 2002. Appraising and Assessing Reflection in Students' Writing on a Structured Worksheet. *Medical Education*, 36, 575-585.
- Pelgrim, E. & Kramer, A. 2013. How Can Medical Education Benefit from the Evidence on Learners Seeking and Using Feedback? *Medical Education*, 47, 225-227.
- Pelgrim, E., Kramer, A., Mokkink, H., van den Elsen, L., Grol, R. & van der Vleuten, C. 2011. In-Training Assessment Using Direct Observation of Single-Patient Encounters: A Literature Review. *Advances in Health Sciences Education*, 16, 131-142.
- Pelgrim, E., Kramer, A., Mokkink, H. & van der Vleuten, C. 2012a. The Process of Feedback in Workplace-Based Assessment: Organisation, Delivery, Continuity. *Medical Education*, 46, 604-612.
- Pelgrim, E., Kramer, A., Mokkink, H. & van der Vleuten, C. 2012b. Quality of Written Narrative Feedback and Reflection in a Modified Mini-Clinical Evaluation Exercise: An Observational Study. *BMC Medical Education*, 12, 1-6.
- Pelgrim, E., Kramer, A., Mokkink, H. & van der Vleuten, C. 2014. Factors Influencing Trainers' Feedback-Giving Behavior: A Cross-Sectional Survey. *BMC Medical Education*, 14, 1-16.
- Pendleton, D., Schofield, T., Tate, P. & Havelock, P. 2003. *The New Consultation: Developing Doctor-Patient Communication*, Oxford, UK, Oxford University Press.
- Pereira, E. & Dean, B. 2009. British Surgeons' Experiences of Mandatory Online Workplace-Based Assessment. *Journal of the Royal Society of Medicine*, 102, 287-293.
- Phan, H. 2008. Predicting Change in Epistemological Beliefs, Reflective Thinking and Learning Styles: A Longitudinal Study. *British Journal of Educational Psychology*, 78, 75-93.
- Phan, H. 2009. Exploring Students' Reflective Thinking Practice, Deep Processing Strategies, Effort, and Achievement Goal Orientations. *Educational Psychology*, 29, 297-313.
- Piaget, J. 1952. *The Origins of Intelligence in Children*, New York, USA, International University Press.
- Planas Lladó, A., Soley, L., Fraguell Sansbelló, R., Pujolras, G., Planella, J., Roura-

- Pascual, N., Suñol Martínez, J. & Moreno, L. 2013. Student Perceptions of Peer Assessment: An Interdisciplinary Study. *Assessment & Evaluation in Higher Education*, 39, 592-610.
- Plasschaert, A., Manogue, M., Lindh, C., McLoughlin, J., Murtomaa, H., Nattestad, A. & Sanz, M. 2007. Curriculum Content, Structure and Ects for European Dental Schools. Part II: Methods of Learning and Teaching, Assessment Procedures and Performance Criteria. *European Journal of Dental Education*, 11, 125-136.
- Plato 1952. *The Dialogues of Plato [Translated by Benjamin Jowett]*, Chicago, London, Toronto, William Benton, Publisher. Encyclopaedia Britannica Inc.
- Playford, D., Kirke, A., Maley, M. & Worthington, R. 2013. Longitudinal Assessment in an Undergraduate Longitudinal Integrated Clerkship: The Mini Clinical Evaluation Exercise (MceX) Profile. *Medical Teacher*, 35, e1416-e1421.
- Pond, K. & ul-Haq, R. 1997. Learning to Assess Students Using Peer Review. *Studies in Educational Evaluation*, 23, 331-348.
- Pope, C. & Mays, N. 2013. *Qualitative Research in Health Care*, Chichester, UK, Wiley.
- Postma, T. & White, J. 2014. Developing Clinical Reasoning in the Classroom – Analysis of the 4c/Id-Model. *European Journal of Dental Education*, Early Online.
- Powell, J. 1989. The Reflective Practitioner in Nursing. *Journal of Advanced Nursing*, 14, 824-832.
- Prescott-Clements, L., van der Vleuten, C., Schuwirth, L., Gibb, E., Hurst, Y. & Rennie, J. 2011. Measuring the Development of Insight by Dental Health Professionals in Training Using Workplace-Based Assessment. *European Journal of Dental Education*, 15, 159-164.
- Prescott-Clements, L., van der Vleuten, C., Schuwirth, L., Hurst, Y. & Rennie, J. 2008. Evidence for Validity within Workplace Assessment: The Longitudinal Evaluation of Performance (Lep). *Medical Education*, 42, 488-95.
- Prescott, L., Norcini, J., McKinlay, P. & Rennie, J. 2002. Facing the Challenges of Competency-Based Assessment of Postgraduate Dental Training: Longitudinal Evaluation of Performance (Lep). *Medical Education*, 36, 92-97.
- Price, M., Handley, K. & Millar, J. 2011. Feedback: Focusing Attention on Engagement. *Studies in Higher Education*, 36, 879-896.
- Price, M., Handley, K., Millar, J. & O'Donovan, B. 2010. Feedback: All That Effort, but What Is the Effect? *Assessment & Evaluation in Higher Education*, 35, 277-289.
- Pritchard, A. 2013. *Ways of Learning: Learning Theories and Learning Styles in the*

- Classroom: Learning Theories and Learning Styles in the Classroom*, Oxon, UK, Taylor & Francis.
- Quantrill, S. & Tun, J. 2012. Workplace-Based Assessment as an Educational Tool. Guide Supplement 31.5 – Viewpoint1. *Medical Teacher*, 34, 417-418.
- Ram, P., Grol, R., Rethans, J. J., Schouten, B., van der Vleuten, C. & Kester, A. 1999. Assessment of General Practitioners by Video Observation of Communicative and Medical Performance in Daily Practice: Issues of Validity, Reliability and Feasibility. *Medical Education*, 33, 447-454.
- Ramani, S. & Krackov, S. 2012. Twelve Tips for Giving Feedback Effectively in the Clinical Environment. *Medical Teacher*, 34, 787-791.
- Ramaprasad, A. 1983. On the Definition of Feedback. *Behavioral Science*, 28, 4-13.
- Ramsden, P. 2003. *Learning to Teach in Higher Education*, Oxon, UK, RoutledgeFalmer.
- Regehr, G. & Norman, G. 1996. Issues in Cognitive Psychology: Implications for Professional Education. *Academic Medicine*, 71, 988-1001.
- Rethans, J., Norcini, J., Barón-Maldonado, M., Blackmore, D., Jolly, B., LaDuca, T., Lew, S., Page, G. & Southgate, L. 2002. The Relationship between Competence and Performance: Implications for Assessing Practice Performance. *Medical Education*, 36, 901-909.
- Roediger, H. & Karpicke, J. 2006. The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, 1, 181-210.
- Roff, S., Chandratilake, M., Mcaleer, S. & Gibson, J. 2011. Preliminary Benchmarking of Appropriate Sanctions for Lapses in Undergraduate Professionalism in the Health Professions. *Medical Teacher*, 33, 234-238.
- Rolfe, I. & McPherson, J. 1995. Formative Assessment: How Am I Doing? *The Lancet*, 345, 837-839.
- Ruth, D. 2013. Teaching Strategy: Reflections on Professional Practice. *Teaching in Higher Education*, 19, 254-265.
- Sabey, A. & Harris, M. 2011. Training in Hospitals: What Do Gp Specialist Trainees Think of Workplacebased Assessments? *Education for Primary Care*, 22, 90-99.
- Sadler, D. 1983. Evaluation and the Improvement of Academic Learning. *The Journal of Higher Education*, 54, 60-79.
- Sadler, D. 1989. Formative Assessment and the Design of Instructional Systems. *Instructional Science*, 18, 119-144.

- Sadler, D. 2005. Interpretations of Criteria-Based Assessment and Grading in Higher Education. *Assessment & Evaluation in Higher Education*, 30, 175-194.
- Sadler, D. 2010. Beyond Feedback: Developing Student Capability in Complex Appraisal. *Assessment & Evaluation in Higher Education*, 35, 535-550.
- Saedon, H., Saedon, M. & Aggarwal, S. 2010. Workplace-Based Assessment as an Educational Tool: Guide Supplement 31.3 – Viewpoint. *Medical Teacher*, 32, e369-e372.
- Sambell, K., McDowell, L. & Brown, S. 1997. “But Is It Fair?”: An Exploratory Study of Student Perceptions of the Consequential Validity of Assessment. *Studies in Educational Evaluation*, 23, 349-371.
- Sanders, J. 2009. The Use of Reflection in Medical Education: Amee Guide No. 44. *Medical Teacher*, 31, 685-695.
- Sanders, J. & Homer, M. 2008. Reflective Learning and the Net Generation. *Medical Teacher*, 30, 877-9.
- Sandywell, B. 2013. *Reflexivity and the Crisis of Western Reason: Logological Investigations: Volume One*, London, UK, Taylor & Francis.
- Sargeant, J., Mann, K., van der Vleuten, C. & Metsemakers, J. 2008. “Directed” Self-Assessment: Practice and Feedback within a Social Context. *Journal of Continuing Education in the Health Professions*, 28, 47-54.
- Sargeant, J., Mann, K., van der Vleuten, C. & Metsemakers, J. 2009. Reflection: A Link between Receiving and Using Assessment Feedback. *Advances in Health Sciences Education*, 14, 399-410.
- Sargeant, J., McNaughton, E., Mercer, S., Murphy, D., Sullivan, P. & Bruce, D. 2011. Providing Feedback: Exploring a Model (Emotion, Content, Outcomes) for Facilitating Multisource Feedback. *Medical Teacher*, 33, 744-749.
- Satterthwaite, J. & Grey, N. 2008. Peer-Group Assessment of Pre-Clinical Operative Skills in Restorative Dentistry and Comparison with Experienced Assessors. *European Journal of Dental Education*, 12, 99-102.
- Schmidt, R. & Lee, T. 2011. *Motor Control and Learning: A Behavioral Emphasis*, Champaign, Illinois, USA, Human Kinetics.
- Schoenrock-Adema, J., Heijne-Penninga, M., van Duijn, M., Geertsma, J. & Cohen-Schotanus, J. 2007. Assessment of Professional Behaviour in Undergraduate Medical Education: Peer Assessment Enhances Performance. *Medical Education*, 41, 836-842.
- Schön, D. 1983. *The Reflective Practitioner: How Professionals Think in Action*, New

- York, USA, Basic Books.
- Schön, D. 1987. *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*, San Francisco, CA, USA, Jossey-Bass Publishers.
- Schoonheim-Klein, M., Habets, L., Aartman, I., van der Vleuten, C., Hoogstraten, J. & van der Velden, U. 2006. Implementing an Objective Structured Clinical Examination (OSCE) in Dental Education: Effects on Students' Learning Strategies. *European Journal of Dental Education*, 10, 226-235.
- Schoonheim-Klein, M., Muijtjens, A., Habets, L., Manogue, M., van der Vleuten, C., Hoogstraten, J. & van der Velden, U. 2008. On the Reliability of a Dental OSCE, Using Sem: Effect of Different Days. *European Journal of Dental Education*, 12, 131-137.
- Schoonheim-Klein, M., Muijtjens, A., Habets, L., Manogue, M., van der Vleuten, C. & van der Velden, U. 2009. Who Will Pass the Dental OSCE? Comparison of the Angoff and the Borderline Regression Standard Setting Methods. *European Journal of Dental Education*, 13, 162-171.
- Schoonheim-Klein, M., Walmsley, A., Habets, L., van der Velden, U. & Manogue, M. 2005. An Implementation Strategy for Introducing an OSCE into a Dental School. *European Journal of Dental Education*, 9, 143-149.
- Schuwirth, L. 2013. Making the Horse Drink: Use of Mini-Cex in an Assessment for Learning View. *Advances in Health Sciences Education*, 18, 1-4.
- Schuwirth, L., Colliver, J., Gruppen, L., Kreiter, C., Mennin, S., Onishi, H., Pangaro, L., Ringsted, C., Swanson, D., van der Vleuten, C. & Wagner-Menghin, M. 2011. Research in Assessment: Consensus Statement and Recommendations from the Ottawa 2010 Conference. *Medical Teacher*, 33, 224-233.
- Schuwirth, L. & van der Vleuten, C. 2010. How to Design a Useful Test: The Principles of Assessment. In: SWANWICK, T. (ed.) *Understanding Medical Education: Evidence, Theory and Practice*. 1st ed. West Sussex, UK: Wiley-Blackwell.
- Schuwirth, L. & van der Vleuten, C. 2011a. General Overview of the Theories Used in Assessment: Amee Guide No. 57. *Medical Teacher*, 33, 783-797.
- Schuwirth, L. & van der Vleuten, C. 2011b. Programmatic Assessment: From Assessment of Learning to Assessment for Learning. *Medical Teacher*, 33, 478-485.
- Setna, Z., Jha, V., Boursicot, K. & Roberts, T. 2010. Evaluating the Utility of Workplace-Based Assessment Tools for Speciality Training. *Best Practice & Research Clinical Obstetrics and Gynaecology*, 24, 767-782.

- Shapiro, J. & Talbot, Y. 1991. Applying the Concept of the Reflective Practitioner to Understanding and Teaching Family Medicine. *Family Medicine*, 23, 450-456.
- Shue, C., Arnold, L. & Stern, D. 2005. Maximizing Participation in Peer Assessment of Professionalism: The Students Speak. *Academic Medicine*, 80, S1-S5.
- Shumway, J. & Harden, R. 2003. Amee Guide No. 25: The Assessment of Learning Outcomes for the Competent and Reflective Physician. *Medical Teacher*, 25, 569-584.
- Shute, V. J. 2008. Focus on Formative Feedback. *Review of Educational Research*, 78, 153-189.
- Simpson, M. 1972. *Medical Education: A Critical Approach*, London, UK, Butterworth Limited.
- Skinner, B. 1954. The Science of Learning and the Art of Teaching. *Harvard Educational Review*, 24, 86-97.
- Skinner, B. 1958. Reinforcement Today. *American Psychologist*, 13, 94-99.
- Skipper, Y. & Douglas, K. 2012. Is No Praise Good Praise? Effects of Positive Feedback on Children's and University Students' Responses to Subsequent Failures. *British Journal of Educational Psychology*, 82, 327-339.
- Sluijsmans, D., Brand-Gruwel, S. & van Merriënboer, J. 2002. Peer Assessment Training in Teacher Education: Effects on Performance and Perceptions. *Assessment & Evaluation in Higher Education*, 27, 443-454.
- Sluijsmans, D., Moerkerke, G., van Merriënboer, J. & Dochy, F. 2001. Peer Assessment in Problem Based Learning. *Studies in Educational Evaluation*, 27, 153-173.
- Smither, J., London, M. & Reilly, R. 2005. Does Performance Improve Following Multisource Feedback? A Theoretical Model, Meta-Analysis, and Review of Empirical Findings. *Personnel Psychology*, 58, 33-66.
- Sobral, D. 2000. An Appraisal of Medical Students' Reflection-in-Learning. *Medical Education*, 34, 182-187.
- Sobral, D. 2001. Medical Students' Reflection in Learning in Relation to Approaches to Study and Academic Achievement. *Medical Teacher*, 23, 508-513.
- Sobral, D. 2005. Medical Students' Mindset for Reflective Learning: A Revalidation Study of the Reflection-in-Learning Scale. *Advances in Health Sciences Education*, 10, 303-314.
- Southgate, L. & van der Vleuten, C. 2014. A Conversation About the Role of Medical Regulators. *Medical Education*, 48, 215-218.
- Spencer, J. 2003. Learning and Teaching in the Clinical Environment. *British Medical*

- Journal*, 326, 591-594.
- Speyer, R., Pilz, W., Van Der Kruis, J. & Brunings, J. 2011. Reliability and Validity of Student Peer Assessment in Medical Education: A Systematic Review. *Medical Teacher*, 33, e572-e585.
- Spronken-Smith, R. & Walker, R. 2010. Can Inquiry-Based Learning Strengthen the Links between Teaching and Disciplinary Research? *Studies in Higher Education*, 35, 723-740.
- Stegeman, J., Schoten, E. & Terpstra, O. 2013. Knowing and Acting in the Clinical Workplace: Trainees' Perspectives on Modelling and Feedback. *Advances in Health Sciences Education*, 18, 597-615.
- Steinert, Y. 2011. Developing Medical Educators: A Journey, Not a Destination. In: SWANWICK, T. (ed.) *Understanding Medical Education: Evidence, Theory and Practice*. 1st ed. West Sussex, UK: Wiley-Blackwell.
- Stenhouse, L. 1975. *An Introduction to Curriculum Research and Development*, London, UK, Heinemann.
- Stern, D. & Papadakis, M. 2006. The Developing Physician — Becoming a Professional. *New England Journal of Medicine*, 355, 1794-1799.
- Stocker, M., Burmester, M. & Allen, M. 2014. Optimisation of Simulated Team Training through the Application of Learning Theories: A Debate for a Conceptual Framework. *BMC Medical Education*, 14, 1-9.
- Strachan, I. & Wilcox, S. 1996. Peer and Self Assessment of Group Work: Developing an Effective Response to Increased Enrolment in a Third-Year Course in Microclimatology. *Journal of Geography in Higher Education*, 20, 343-353.
- Streiner, D. & Norman, G. 2008. *Health Measurement Scales – a Practical Guide to Their Development and Use*, Oxford, UK, Oxford University Press.
- Subramanian, J., Anderson, V., Morgaine, K. & Thomson, W. 2013. Improving the Quality of Educational Strategies in Postgraduate Dental Education Using Student and Graduate Feedback: Findings from a Qualitative Study in New Zealand. *European Journal of Dental Education*, 17, e151-e158.
- Sumsion, J. & Fleet, A. 1996. Reflection: Can We Assess It? Should We Assess It? *Assessment & Evaluation in Higher Education*, 21, 121-130.
- Swanwick, T. 2005. Informal Learning in Postgraduate Medical Education: From Cognitivism to 'Culturism'. *Medical Education*, 39, 859-865.
- Swanwick, T. 2011. *Understanding Medical Education: Evidence, Theory and Practice*, West Sussex, UK, Wiley-Blackwell.

- Sweet, J., Wilson, J. & Pugsley, L. 2008. Chairside Teaching and the Perceptions of Dental Teachers in the UK. *British Dental Journal*, 205, 565-569.
- Sweller, J., van Merriënboer, J. & Paas, F. 1998. Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10, 251-296.
- Tavakol, M. & Dennick, R. 2011. Post-Examination Analysis of Objective Tests. *Medical Teacher*, 33, 447-58.
- Tavakol, M. & Dennick, R. 2012. Post-Examination Interpretation of Objective Test Data: Monitoring and Improving the Quality of High-Stakes Examinations: Amee Guide No. 66. *Medical Teacher*, 34, e161-e175.
- Tavakol, M. & Dennick, R. 2013. Psychometric Evaluation of a Knowledge Based Examination Using Rasch Analysis: An Illustrative Guide: Amee Guide No. 72. *Medical Teacher*, 35, e838-e848.
- Taylor, C., Grey, N. & Satterthwaite, J. 2013. A Comparison of Grades Awarded by Peer Assessment, Faculty and a Digital Scanning Device in a Pre-Clinical Operative Skills Course. *European Journal of Dental Education*, 17, e16-e21.
- Taylor, D. & Hamdy, H. 2013. Adult Learning Theories: Implications for Learning and Teaching in Medical Education: Amee Guide No. 83. *Medical Teacher*, 35, e1561–e1572.
- Teich, S., Demko, C. & Lang, L. 2014. Students' Perception of Peer-Assessment in the Context of a Treatment Planning Course. *European Journal of Dental Education*, Early Online.
- ten Cate, O. & Durning, S. 2007. Dimensions and Psychology of Peer Teaching in Medical Education. *Medical Teacher*, 29, 546-552.
- ten Cate, O., Snell, L., Mann, K. & Vermunt, J. 2004. Orienting Teaching toward the Learning Process. *Academic Medicine*, 79, 219-228.
- Teunissen, P., Stapel, D., van der Vleuten, C., Scherpbier, A., Boor, K. & Scheele, F. 2009. Who Wants Feedback? An Investigation of the Variables Influencing Residents' Feedback-Seeking Behavior in Relation to Night Shifts. *Academic Medicine*, 84, 910-917.
- The Royal College of Surgeons of England. 2014. *Intercollegiate Surgical Curriculum Programme (Iscp)*, [Online]. Available: <https://www.iscp.ac.uk/dental/wbas.aspx> [Accessed June 2014].
- Thorndike, E. 1911. *Animal Intelligence: Experimental Studies*, New York, USA, Macmillan.
- Tolsgaard, M., Arendrup, H., Pedersen, P. & Ringsted, C. 2013. Feasibility of Self-

- Directed Learning in Clerkships. *Medical Teacher*, 35, e1409–e1415.
- Tomás, I., Millán, U., Casares, M., Abad, M., Ceballos, L., Gómez-Moreno, G., Hidalgo, J., Llena, C., López-Jornet, P., Machuca, M., Monticelli, F. & Palés, J. 2013. Analysis of the 'Educational Climate' in Spanish Public Schools of Dentistry Using the Dundee Ready Education Environment Measure: A Multicenter Study. *European Journal of Dental Education*, 17, 159-168.
- Tonni, I. & Oliver, R. 2013a. Acceptability of a Reflective E-Portfolio Instituted in an Orthodontic Specialist Programme: A Pilot Study. *European Journal of Dental Education*, 17, 177-184.
- Tonni, I. & Oliver, R. 2013b. A Delphi Approach to Define Learning Outcomes and Assessment. *European Journal of Dental Education*, 17, e173-e180.
- Topping, K. 1996. The Effectiveness of Peer Tutoring in Further and Higher Education: A Typology and Review of the Literature. *Higher Education*, 32, 321-345.
- Topping, K. 1998. Peer Assessment between Students in Colleges and Universities. *Review of Educational Research*, 68, 249-276.
- Topping, K. 2005. Trends in Peer Learning. *Educational Psychology*, 25, 631-645.
- Torre, D., Simpson, D., Elnicki, D., Sebastian, J. & Holmboe, E. 2007. Feasibility, Reliability and User Satisfaction with a Pda-Based Mini-Cex to Evaluate the Clinical Skills of Third-Year Medical Students. *Teaching and Learning in Medicine*, 19, 271-277.
- Tsang, A. & Walsh, L. 2010. Oral Health Students' Perceptions of Clinical Reflective Learning – Relevance to Their Development as Evolving Professionals. *European Journal of Dental Education*, 14, 99-105.
- UK Government "Dentist Act 1984". <http://www.legislation.gov.uk/ukpga/1984/24>.
- Unistats. 2014. *(Nss) Student Satisfaction: Assessment and Feedback* [Online]. UK: Unistats. Available: <http://unistats.direct.gov.uk/subjects/satisfaction/10003645-UBDS5NSDN> [Accessed May 2014].
- Urquhart, L., Rees, C. & Ker, J. 2014. Making Sense of Feedback Experiences: A Multi-School Study of Medical Students' Narratives. *Medical Education*, 48, 189-203.
- van de Ridder, J., Stokking, K., McGaghie, W. & ten Cate, O. 2008. What Is Feedback in Clinical Education? *Medical Education*, 42, 189-197.
- van den Boom, G., Paas, F. & van Merriënboer, J. 2007. Effects of Elicited Reflections Combined with Tutor or Peer Feedback on Self-Regulated Learning and Learning Outcomes. *Learning and Instruction*, 17, 532-548.

- van der Leeuw, R. & Slootweg, I. 2013. Twelve Tips for Making the Best Use of Feedback. *Medical Teacher*, 35, 348-351.
- van der Vleuten, C. 1996. The Assessment of Professional Competence: Developments, Research and Practical Implications. *Advances in Health Sciences Education*, 1, 41-67.
- van der Vleuten, C. & Schuwirth, L. 2005. Assessing Professional Competence: From Methods to Programmes. *Medical Education*, 39, 309-317.
- van der Vleuten, C., Schuwirth, L., Driessen, E., Dijkstra, J., Tigelaar, D., Baartman, L. & van Tartwijk, J. 2012. A Model for Programmatic Assessment Fit for Purpose. *Medical Teacher*, 34, 205-214.
- van der Zwet, J., Zwietering, P., Teunissen, P., van der Vleuten, C. & Scherpbier, A. 2011. Workplace Learning from a Socio-Cultural Perspective: Creating Developmental Space During the General Practice Clerkship. *Advances in Health Sciences Education*, 16, 359-373.
- Van Dijk, D. & Kluger, A. 2011. Task Type as a Moderator of Positive/Negative Feedback Effects on Motivation and Performance: A Regulatory Focus Perspective. *Journal of Organizational Behavior*, 32, 1084-1105.
- van Manen, M. 1995. On the Epistemology of Reflective Practice. *Teachers and Teaching*, 1, 33-50.
- Van Rosendaal, G. & Jennett, P. 1992. Resistance to Peer Evaluation in an Internal Medicine Residency. *Academic Medicine*, 67, 63.
- Veloski, J., Boex, J., Grasberger, M., Evans, A. & Wolfson, D. 2006. Systematic Review of the Literature on Assessment, Feedback and Physicians' Clinical Performance*: Beme Guide No. 7. *Medical Teacher*, 28, 117-128.
- Vivekananda-Schmidt, P., Marshall, M., Stark, P., McKendree, J., Sandars, J. & Smithson, S. 2011. Lessons from Medical Students' Perceptions of Learning Reflective Skills: A Multi-Institutional Study. *Medical Teacher*, 33, 846-850.
- Vygotsky, L. 1978. *Mind in Society: The Development of Higher Psychological Processes*, Cambridge, MA, Harvard University Press.
- Wald, H., Borkan, J., Taylor, J., Anthony, D. & Reis, S. 2012. Fostering and Evaluating Reflective Capacity in Medical Education: Developing the Reflect Rubric for Assessing Reflective Writing. *Academic Medicine*, 87, 41-50.
- Wald, H., Davis, S., Reis, S., Monroe, A. & Borkan, J. 2009. Reflecting on Reflections: Enhancement of Medical Education Curriculum with Structured Field Notes and Guided Feedback. *Academic Medicine*, 84, 830-837.

- Walker, M. 2014. The Quality of Written Peer Feedback on Undergraduates' Draft Answers to an Assignment, and the Use Made of the Feedback. *Assessment & Evaluation in Higher Education*, Early Online.
- Wallace, J., Blinkhorn, A. & Blinkhorn, F. 2013. Reflective Folios for Dental Hygiene Students: What Do They Tell Us About a Residential Aged Care Student Placement Experience? *European Journal of Dental Education*, 17, 236-240.
- Wallman, A., Lindblad, A., Hall, S., Lundmark, A. & Ring, L. 2008. A Categorization Scheme for Assessing Pharmacy Students' Levels of Reflection During Internships. *American Journal of Pharmaceutical Education*, 72, 1-10.
- Walsh, C., Ling, S., Wang, C. & Carnahan, H. 2009. Concurrent Versus Terminal Feedback: It May Be Better to Wait. *Academic Medicine*, 84, S54-S57.
- Wang, V. 2012. Understanding and Promoting Learning Theories. *International Forum of Teaching & Studies*, 8, 5-11.
- Wang, V. & King, K. 2007. Confucius and Mezirow—Understanding Mezirow's Theory of Reflectivity from Confucian Perspectives: A Model and Perspective. *Comparative Adult Education Around the Globe*, 8, 253-275.
- Wass, V. 2011. Skill-Based Assessment. In: CANTILLON, P. & WOOD, D. (eds.) *Abc of Learning and Teaching in Medicine*. West Sussex, UK: Wiley-Blackwell.
- Wass, V., van der Vleuten, C., Shatzer, J. & Jones, R. 2001. Assessment of Clinical Competence. *The Lancet*, 357, 945-949.
- Watling, C. 2014. Unfulfilled Promise, Untapped Potential: Feedback at the Crossroads. *Medical Teacher*, Early Online.
- Watling, C., Driessen, E., van der Vleuten, C. & Lingard, L. 2012a. Learning from Clinical Work: The Roles of Learning Cues and Credibility Judgements. *Medical Education*, 46, 192-200.
- Watling, C., Driessen, E., van der Vleuten, C., Vanstone, M. & Lingard, L. 2012b. Understanding Responses to Feedback: The Potential and Limitations of Regulatory Focus Theory. *Medical Education*, 46, 593-603.
- Watling, C., Driessen, E., van der Vleuten, C., Vanstone, M. & Lingard, L. 2013. Music Lessons: Revealing Medicine's Learning Culture through a Comparison with That of Music. *Medical Education*, 47, 842-850.
- Watling, C., Kenyon, C., Zibrowski, E., Schulz, V., Goldszmidt, M., Singh, I., Maddocks, H. & Lingard, L. 2008. Rules of Engagement: Residents' Perceptions of the in-Training Evaluation Process. *Academic Medicine*, 83, S97-S100.
- Watson, J. 1913. Psychology as the Behaviorist Views It. *Psychological Review*, 20,

- 158-177.
- Watson, M., Wong, D., Kluger, R., Chuan, A., Herrick, M., Ng, I., Castanelli, D., Lin, L., Lansdown, A. & Barrington, M. 2014. Psychometric Evaluation of a Direct Observation of Procedural Skills Assessment Tool for Ultrasound-Guided Regional Anaesthesia. *Anaesthesia*, 69, 604-612.
- Weaver, M. 2006. Do Students Value Feedback? Student Perceptions of Tutors' Written Responses. *Assessment & Evaluation in Higher Education*, 31, 379-394.
- Wenger, E. 1998. *Communities of Practice: Learning, Meaning, and Identity*, Cambridge, UK, Cambridge University Press.
- Weston, P. & Smith, C. 2014. The Use of Mini-Cex in UK Foundation Training Six Years Following Its Introduction: Lessons Still to Be Learned and the Benefit of Formal Teaching Regarding Its Utility. *Medical Teacher*, 36, 155-163.
- Wetmore, A., Boyd, L., Bowen, D. & Pattillo, R. 2010. Reflective Blogs in Clinical Education to Promote Critical Thinking in Dental Hygiene Students. *Journal of Dental Education*, 74, 1337-1350.
- White, J. & Sharma, N. 2012. "Who Writes What?" Using Written Comments in Team-Based Assessment to Better Understand Medical Student Performance: A Mixed-Methods Study. *BMC Medical Education*, 12, 1-22.
- Wiecha, J., Vanderschmidt, H. & Schilling, K. 2002. Heal: An Instructional Design Model Applied to an Online Clerkship in Family Medicine. *Academic Medicine*, 77, 925-926.
- Wiggins, G. 1993. *Assessing Student Performance: Exploring the Purpose and Limits of Testing*, West Sussex, UK, Wiley.
- Wilkinson, J., Crossley, J., Wragg, A., Mills, P., Cowan, G. & Wade, W. 2008. Implementing Workplace-Based Assessment across the Medical Specialties in the United Kingdom. *Medical Education*, 42, 364-373.
- Williams, E. 1992. Student Attitudes Towards Approaches to Learning and Assessment. *Assessment & Evaluation in Higher Education*, 17, 45-58.
- Williams, R., Klamen, D. & McGaghie, W. 2003. Cognitive, Social and Environmental Sources of Bias in Clinical Performance Ratings. *Teaching and Learning in Medicine*, 15, 270-292.
- Williams, R., Wessel, J., Gemus, M. & Foster-Seargeant, E. 2002. Journal Writing to Promote Reflection by Physical Therapy Students During Clinical Placements. *Physiotherapy Theory and Practice*, 18, 5-15.
- Wingate, U. 2010. The Impact of Formative Feedback on the Development of

- Academic Writing. *Assessment & Evaluation in Higher Education*, 35, 519-533.
- Wojtczak, A. 2002. Glossary of Medical Education Terms: Part 1. *Medical Teacher*, 24, 216-219.
- Wong, F., Kember, D., Chung, L. & Yan, L. 1995. Assessing the Level of Student Reflection from Reflective Journals. *Journal of Advanced Nursing*, 22, 48-57.
- Wood, D. 2011. Formative Assessment. In: SWANWICK, T. (ed.) *Understanding Medical Education: Evidence, Theory and Practice*. West Sussex, UK: Wiley-Blackwell.
- Wood, D., Bruner, J. & Ross, G. 1976. The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry*, 17, 89-100.
- Woodman, T., Pee, B., Fry, H. & Davenport, E. 2002. Practice-Based Learning: Emerging Professional Characteristics, Self-Concepts and Patterns of Knowing in Dental Training. *European Journal of Dental Education*, 6, 9-15.
- Wragg, A., Wade, W., Fuller, G., Cowan, G. & Mills, P. 2003. Assessing the Performance of Specialist Registrars. *Clinical Medicine*, 3, 131-134.
- Wright, C., Richards, S., Hill, J., Roberts, M., Norman, G., Greco, M., Taylor, M. & Campbell, J. 2012. Multisource Feedback in Evaluating the Performance of Doctors: The Example of the UK General Medical Council Patient and Colleague Questionnaires. *Academic Medicine*, 87, 1668-1678.
- Yardley, S., Teunissen, P. & Dornan, T. 2012. Experiential Learning: Amee Guide No. 63. *Medical Teacher*, 34, e102-e115.